DOI:
10.1016/j.fsigen.2015.06.004

Document Version
Peer reviewed version

*Title Page (with Author Details)

Forensic ancestry analysis with two capillary electrophoresis ancestry informative marker (AIM) panels: Results of a collaborative EDNAP exercise

C. Santos[a,1], M. Fondevila[a,1], D. Ballard[b,1], R. Banemann[c], A.M. Bento[d], C. Børsting[e,1], W. Branicki[f,2], F. Brisighelli[g], M. Burrington[h], T. Capal[i], L. Chaitanya[j], R. Daniel[k], V. Decroyer[l], R. England[m], K.B. Gettings[n], T.E. Gross[o,1], C. Haas[p], J. Harteveld[q], P. Hoff-Olsen[r], A. Hoffmann[c], M. Kayser[j], P. Kohler[r,2], A. Linacre[s], M. Mayr-Eduardoff[t,1], C. McGovern[m], N. Morling[e,1 l], G. O'Donnell[h], W. Parson[t,u,1], V.L. Pascali[g], M.J. Porto[d], A. Roseth[r], P.M. Schneider[o,1], T. Sijen[q], V. Stenzl[i], D. Syndercombe Court[b,1], J.E. Templeton[s], M. Turanska[v], P.M. Vallone[n], R.A.H. van Oorschot[k], L. Zatkalikova[v], The EUROFORGEN- NoE Consortium; Á. Carracedo[a,1], C. Phillips[a,1]*

[a] Forensic Genetics Unit, Institute of Forensic Sciences, University of Santiago de Compostela, Santiago de Compostela, Spain

[b] Department of Forensic and Analytical Science, Faculty of Life Science, King's College London, UK

[c] Federal Criminal Police Office, Wiesbaden, Germany.

[d] Forensic Genetic and Biology Service, Centre Branch, National Institute of Legal Medicine and Forensic Sciences , Coimbra, Portugal

[e] Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Frederik V's Vej 11, Copenhagen, Denmark

[f] Section of Forensic Genetics, Institute of Forensic Research, Kraków, Poland

[g] Forensic Genetics Laboratory, Institute of Legal Medicine, Università Cattolica del Sacro Cuore, Rome, Italy

[h] Forensic Science Laboratory, Dublin, Ireland

[i] Department of Forensic Genetics, Institute of Criminalistics, Prague, Czech Republic

[j] Department of Forensic Molecular Biology, Erasmus MC University Medical Centre Rotterdam, Rotterdam, The Netherlands

[k] Office of the Chief Forensic Scientist, Forensic Services Department, Victoria Police, Australia

[l] National Institute of Criminalistics and Criminology, Chaussée de Vilvoorde 100, Brussels, Belgium

[m] ESR, Private Bag 92021, Auckland, New Zealand

[n] Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA

[o] Institute of Legal Medicine, Faculty of Medicine, University of Cologne, Cologne, Germany

[p] Zurich Institute of Forensic Medicine, University of Zurich, Zurich, Switzerland

[q] Department of Human Biological Traces, Netherlands Forensic Institute, The Hague, The Netherlands

[r] Department of Forensic Biology, Norwegian Institute of Public Health, Oslo, Norway

[s] School of Biological Sciences, Flinders University, Adelaide, South Australia 5042, Australia

[t] Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck, Austria

[u] Forensic Science Program, The Pennsylvania State University, University Park, PA, USA

[v] Institute of Forensic Science, Ministry of the Interior, Department of Biology and DNA Analysis, Slovenská Lupca, Slovakia

[1] Also participated in the EUROFORGEN-NoE inter-laboratory ancestry analysis pilot exercise

[2] Participated in the EUROFORGEN-NoE pilot exercise only

* Corresponding author. E-mail address: c.phillips@mac.com (C. Phillips).

- Nineteen laboratories completed a collaborative EDNAP exercise to evaluate two forensic ancestry informative marker (AIM) assays and accompanying statistical tools to infer ancestry from the genotype data.

- Laboratories were sent primers, reference data and five test DNAs of undisclosed origin plus an unmarked DNA mixture (but reported to be one of the samples).

- Fourteen laboratories successfully genotyped the DNAs with a 34-plex SNP assay using SNaPshot, achieving 96.1% profile completeness and 93.5% genotype concordance.

- All laboratories successfully genotyped the DNAs with a 46-plex Indel assay using dye-labelled PCR primers, achieving 99.8% profile completeness and genotype concordance.

- All laboratories identified the mixed DNA sample, indicated by disrupted peak height ratios in the Indel profile and three-allele patterns in SNP rs5030240. 18/19 laboratories assigned the correct ancestry to each of the test DNAs of unknown origin, obtaining likelihood ratios from 80 markers in the range: 1.25E+07 to 1.78E+41.

Forensic ancestry analysis with two capillary electrophoresis ancestry informative marker (AIM) panels: Results of a collaborative EDNAP exercise

Abstract

There is increasing interest in forensic ancestry tests, which are part of a growing number of DNA analyses that can enhance routine profiling by obtaining additional genetic information about unidentified DNA donors. Nearly all ancestry tests use single nucleotide polymorphisms (SNPs), but these currently rely on SNaPshot single base extension chemistry that can fail to detect mixed DNA. Insertion-deletion polymorphism (Indel) tests have been developed using dye-labeled primers that allow direct capillary electrophoresis detection of PCR products (PCR-to-CE). PCR-to-CE maintains the direct relationship between input DNA and signal strength as each marker is detected with a single dye, so mixed DNA is more reliably detected. We report the results of a collaborative inter-laboratory exercise of 19 participants (15 from the EDNAP European DNA Profiling group) that assessed a 34-plex SNP test using SNaPshot and a 46-plex Indel test using PCR-to-CE. Laboratories were asked to type five samples with different ancestries and detect an additional mixed DNA sample. Statistical inference of ancestry was made by participants using the *Snipper* online Bayes analysis portal plus an optional PCA module that analyzes the genotype data alongside calculation of Bayes likelihood ratios**.** Exercise results indicated consistent genotyping performance from both tests, reaching a particularly high level of reliability for the Indel test. SNP genotyping gave 93.5% concordance (compared to the organizing laboratory's data) that rose to 97.3% excluding one laboratory with a large number of miscalled genotypes. Indel genotyping gave a higher concordance rate of 99.8% and a reduced no-call rate compared to SNP analysis. All participants detected the mixture from their Indel peak height data and successfully assigned the correct ancestry to the other samples using *Snipper*, with the exception of one laboratory with SNP miscalls that incorrectly assigned ancestry of two samples and did not obtain informative likelihood ratios for a third. Therefore, successful ancestry assignments were achieved by participants in 92 of 95 *Snipper* analyses. This exercise demonstrates that ancestry inference tests based on binary marker sets can be readily adopted by laboratories that already have well-established CE regimes in place. The Indel test proved to be easy to use and allowed all exercise participants to detect the DNA mixture as well as achieving complete and concordant profiles in nearly all cases. Lastly, two participants successfully ran parallel next-generation sequencing analyses (each using different systems) and achieved high levels of genotyping concordance using the exercise PCR primer mixes unmodified.

Keywords: Ancestry; SNPs; Indels; AIMs; Bayes analysis; Principal Component Analysis (PCA)

1. Introduction

DNA-based forensic ancestry tests have the capacity to provide key information about unidentified DNA donors, which can be particularly useful when police investigators do not have reliable eyewitness descriptions or if the STR profiling data fails to give a DNA database match [1]. Therefore, tests for the inference of ancestry can be grouped alongside forensic DNA phenotyping (FDP) tests such as *HIrisPlex* [2] in a growing array of new technologies that have the potential to take forensic DNA analysis well beyond simple identification [3,4]. For such tests to be effective in routine forensic use they must be sensitive; easy to run using validated DNA detection instruments; and, being mainly composed of binary loci, they should have a reasonably robust way to detect mixed DNA so that apparent heterozygotes are not mistyped. In addition, the genetic data obtained must be easy to interpret. Ideally, it should be straightforward to use the genotypes to calculate a set of Bayes likelihoods for particular ancestries (or phenotypes) in comparison to reference populations whose patterns of genetic variation are already well defined. Although STRs can provide a degree of ancestry information [5,6] and Y-chromosome/mtDNA variation is highly differentiated geographically, there are widely discussed reasons why stand-alone autosomal SNP tests provide more reliable indications of a person's ancestry [7-9].

For the last ten years, forensic SNP genotyping has relied on the SNaPshot single base extension system to create relatively large-scale PCR and extension multiplexes followed by capillary electrophoresis (CE) of the dye-labeled products using standard run conditions. In this way, FDP and ancestry analysis tests [2,8-12] have been developed using single-tube amplification reactions that are highly sensitive and use validated CE regimes [10,13]. One drawback of SNP genotyping with SNaPshot is the inability to distinguish the highly skewed heterozygote peaks often seen in normal DNA with this technique, from the imbalanced peaks common to mixtures. This is mainly due to the SNaPshot terminator chemistry using dyes with much stronger blue/green fluorescence (G/A) compared to yellow/red (C/T) [14]. Therefore, despite their widespread use and evident sensitivity, forensic SNaPshot tests can be inefficient in detecting mixtures. Indel tests have been developed in recent years for identification [15-17] and ancestry analysis [18-20] detecting dye-labeled PCR products sent directly to CE from the amplification stage (PCR-to-CE). The benefits of short amplicon lengths and high levels of multiplexing that SNPs provide, are kept with Indel genotyping in this way. However, peak height ratios in heterozygotes are more balanced within any one locus than those of SNaPshot so mixed DNA is more easily detected from the resulting imbalanced patterns [17]. Two CE-based forensic ancestry tests have been established that offer complimentary characteristics: a SNaPshot assay of 34 ancestry informative marker (AIM) SNPs containing some of the most population-differentiated loci (herein 34-plex, [11]) plus a PCR-to-CE assay of 46 AIM-Indels [19] that offers comparable population differentiation to AIM-SNPs, but much greater sensitivity to mixed DNA. This report describes the use of these two assays in an inter-laboratory exercise of 15 participants from the European DNA Profiling (EDNAP) group, and 4 overseas participants, organized by the University of Santiago de Compostela (USC). As a preamble to the EDNAP exercise, the EUROFORGEN-NoE Consortium ran a similar small-scale inter-laboratory exercise to establish the test framework and gauge the transportability of the assay primer sets. As part of the Consortium's networking remit, the primer mixes used for the EDNAP exercise were purchased, optimized and packaged by USC along with test DNAs with known ancestries (undisclosed to participants). These test components are freely available in trial quantities for the forensic community to assess for themselves (available from USC upon request).

The exercise had three main goals: i) for laboratories to assess the relative ease-of-use and reliability of the two assays by genotyping test DNAs, whenever possible, using each participant's own CE regimes; ii)

for laboratories to use the statistical ancestry inference tools developed at USC and part of the *Snipper* data analysis portal [11]; iii) to assess the ability of each assay to detect mixtures by including an unmarked mixed-donor sample amongst the test DNAs. This third goal was analyzed further by assessing the Indel heterozygote peak height balance in normal DNA across the range of participant's laboratory setups, in comparison to peaks in the mixed sample. As well as the 15 European laboratories including USC, two participants were from Australia, one from New Zealand and one from the USA. All but three laboratories had participated in the preceding EDNAP *IrisPlex* exercise that applied SNaPshot analysis to the genotyping of six FDP SNPs [21]. Five EDNAP laboratories, were part of the EUROFORGEN-NoE pilot ancestry exercise.

2. Materials and methods

2.1. Primer sets, test DNA samples and assay protocols

Six quantified DNA samples (10 µl volumes at 0.5 ng/µl) plus primer mixes sufficient for 20 reactions were sent to participants who used their own PCR and SNaPshot reaction components. For the Indel assay, PCRs only required the combination of 2x Qiagen Multiplex PCR Master Mix (Qiagen, Hilden, Germany) with the primer mix and DNA. The SNaPshot PCR and extension primer sets plus the Indel PCR primer mix were prepared as previously described [11,19] and were dispatched with the DNA samples at ambient temperature. Some package transit times outside Europe exceeded one week, but the stability of both SNP and Indel primer sets had been previously assessed for the EUROFORGEN-NoE pilot exercise by carefully testing the profile quality obtained from batches of primers originally sent to the US participant and one in Australia, who were also part of the subsequent exercise.

The test DNAs were given anonymized codes and comprised five volunteer donors, each with a different continental origin of: East Asian, European, Oceanian, Native American or African ancestries. With the geographic distribution of these samples, examples of all alleles in 80 markers were observed when genotyped by USC, except SNP: rs1573020 (all A homozygotes) and Indels: rs35451359 and rs33974167 (all short-allele 'A' homozygotes) plus rs2307998 (all long-allele 'C' homozygotes). In this way, more than 97.6% of component marker alleles could be identifiable in the profiles of the test DNAs. A rare third allele in Indel: rs25584 was found in one test DNA. The sixth test sample was an artificial mixture combining a 1:3 ratio of additional European and East Asian volunteer donors (herein M1 and M3 respectively). Note that 34-plex has two tri-allelic SNPs and one: rs5030240 showed three allele patterns in the mixed DNA sample (other examples in [11]).

Participants were told that one sample was mixed and were asked to identify it, then assign ancestries to the others using Bayes analysis and Principal Component Analysis (PCA) in *Snipper,* as detailed in section 2.2. The above primer volumes were sufficient to allow participants to begin their analyses with the 9947A positive control DNA used in many STR kits.

Protocols for PCR, SNaPshot extension reactions and CE were sent in the form of an Excel laboratory calculator (Supplementary File S1) plus fragment mobility panels-and-bins files (Supplementary Files S2) that formed templates for participants to adapt to their own CE regimes when necessary. The 9947A DNA acted as a universal point of reference for the peak patterns typical of both assays and example electropherograms were provided to participants, as shown in Fig. 1. Although Indel amplified fragments separate well using all POP polymer types, participants were recommended to use POP-4 for 34-plex genotyping as peak positions are less well separated at the low size range using POP-7. Supplementary

Table S1 lists the CE regimes chosen by participants, indicating that most applied a 3130 or 3500 detector with POP-4 (13 and 3 respectively), although two used a 3130 with POP-7 and one successfully typed SNPs with a 3100 and POP-6. Lastly, participants were advised that Indel PCR products could require dilution prior to CE to obtain optimum peak patterns free from excessive signal pull-up.

2.2. Preliminary ancestry checks of test DNAs and use of the Snipper data analysis portal

Although this section reports ancestry analysis results, these analyses were made by USC to evaluate the ancestry of the exercise test samples prior to their dispatch. This process also checked the reference population data supplied and ensured test samples were suitably representative of each of the ancestries the participants were asked to identify.

The *Snipper* portal provides a Bayes classifier accessing population reference data in place in the website, including fixed training sets for three, four or five main continental HGDP-CEPH population groups, for 34 SNPs and/or 46 Indels (these training set genotypes are provided in Supplementary File S3.2). The fixed data options assess one uploaded profile at a time, which is compared to a training set selected by the user. Partial profiles can be uploaded with NN genotypes (or partial genotypes, e.g. 'CN'). Indel data has an identical framework but with 'AC coded' genotypes, where A=short alleles, C=long and is reserved for novel third alleles. Participants were asked to use the fixed training set option in *Snipper* to make ancestry inferences. However, no guidance was given on choice of training set, which influences calculation of the likelihood ratios (herein LRs). For example, selecting a five-group training set for 34-plex SNP data will lead to lower LRs for East Asian assignments as this marker set lacks AIMs sufficiently differentiated to distinguish Oceanians and Native Americans from East Asians. As a rule-of-thumb, 34-plex profiles are optimally analyzed with three-group data (Africa-Europe-East Asia), Indel profiles provide high ancestry assignment LRs for these groups plus Americans, as this differentiation was targeted in their original selection [19]. When combined 80-marker data is used, the differentiation of the fifth Oceanian population group can be accomplished, although Indel data alone can also distinguish Oceanians with minimal error [19].

To check the *Snipper* fixed training sets and test samples used, three ancestry analyses were applied to the genotype data prior to the exercise and results are summarized in Fig. 2. First, the 80-marker reference data was cross-validated with *Snipper* (each training set profile removed and classified by remaining data). The Fig. 2 upper plot shows the distribution of probabilities in ranked order of $\log_{10}$ LR values, i.e. the lowest LR from five population comparisons (data in Supplementary File S3.3). The grey line of LR=1 represents balanced odds, so points below this line show misclassifications. East Asian training set profiles gave five misclassifications, all assigned as American (5/226=2.2% error). However, none of their LRs exceeded 750, so applying a threshold value of 1000 led to error-free East Asian assignments, but a non-classification rate of 3.54% (8/226). Fig. 2 indicates the LRs for test samples, mixture donors and 9947A tend to fall in the middle to upper range of training set LRs in nearly all cases.

In addition to obtaining LRs, it can be helpful to compare patterns of variation in reference population data to samples of unknown geographic origin by applying *STRUCTURE* and PCA. Both provide an intuitive way to make such comparisons [3,22,23] and can be useful to alert the analyst that a forensic sample of unknown origin may be from an admixed individual with co-ancestry. Following review of the EUROFORGEN-NoE ancestry exercise results, a two-dimensional PCA module (plotting the first two principal components or PCs) was developed for *Snipper* that allows analysis of multiple profiles plotted directly onto reference data. The *Snipper* output lists the Bayes analysis data for each profile and their

positions are labeled on a PC1-PC2 PCA plot (no PC3 estimates are currently made). Participants were provided with the input file of training set genotypes and a link to the *Snipper* PCA module to enable graphical analysis of test DNAs and 9947A.

The middle graphics of Fig. 2 show *STRUCTURE* cluster plots (an optimum K=5 genetic clusters inferred from data) matched to the order of training set LRs charted above. The enlarged cluster plots for samples A-E on the right indicate an absence of co-ancestry, i.e. their cluster plots have almost no membership to multiple genetic clusters. Likewise, cluster plots on the left for mixture components M1 and M3 show no multiple cluster membership, whereas sample F has approximately equal joint membership to the relevant clusters. The lower graphics show three PCAs made with *Snipper*, with reference cluster colours matched to the *STRUCTURE* data. PCA plot A is a 3-group analysis of sample F and M1-M3 components. The position of F highlights the fact that population admixture and mixed DNA can give indistinguishable PCA patterns, emphasizing the need to efficiently detect mixed DNA in forensic ancestry analysis. PCA plot B is a 5-group analysis showing samples A-E plus 9947A are distributed into their expected clusters, although in these 2-PC plots the Oceanian and American clusters show some overlap with East Asians. To better differentiate these three groups, a PCA can be made of just three possible groups to obtain a more distinct separation, as shown in PCA plot C analyzing the three test samples from less differentiated population groups.

3. Results

3.1. Genotyping performance of the SNP assay

Supplementary Table S1 summarizes the CE regimes used by participants and indicates five did not pursue SNaPshot genotyping of SNPs but elected to just analyze and report Indel genotypes. Given the complexities of reading electropherograms consisting of 32 peak pairs plus two triple-peak positions, this was considered to be a reasonable decision and Indel data alone was collected from these laboratories. The number of SNaPshot no-calls and miscalls recorded for the five test samples A-E, from 14 participants reporting SNP data, are summarized in Fig. 3A. SNPs are listed in order of decreasing genotyping performance for participants, by ranking loci in increasing miscall rate followed by increasing no-calls. Therefore, rs2065160, rs3785181 and rs8986788 are the most robustly genotyped SNPs in 34-plex, with all 14 laboratories identifying peaks in five samples, although laboratory #17 had one genotype miscall in each SNP. At the other extreme, rs239031 was both the most difficult SNP to genotype and the least reliably genotyped, with laboratories #8 and #21 not assigning genotypes to all or most samples, bringing the overall call rate down to 80%, well below those of the other 33 SNPs. Genotyping concordance for rs239031 was also the lowest, with 81.4% of genotypes correctly called. High no-call rates for certain other SNPs tended to cluster with participants: rs1573020 was not genotyped in laboratory #21; rs881929 in #20; rs1886510 and rs2304925 in #19, despite other laboratories genotyping these SNPs without problems. Only 1 of 5 genotypes was called by laboratory #13 for rs182549. Average SNP call and genotype concordance rates shown at the bottom right of Fig. 3A reached 96.3% and 93.5% respectively. The genotype completeness of ~96% equates to approximately one missing SNP call per 34-plex profile. Laboratory #17 had evident problems recognizing and accurately calling their SNaPshot electropherograms with less than half the successful genotype calls made by the other participants reporting SNPs. Therefore, when considering concordance amongst 13/14 participants, the value rose to 97.3%. Although one other laboratory #6 had slightly below-average genotyping concordance, no obvious connection could be made between the CE regimes used by participants and miscalls seen in certain SNPs. Nevertheless, there are known issues previously recognized at USC in some 34-plex SNPs and several of these were observed in the electropherograms from participants. Certain mobility or non-specific peak patterns can explain a proportion of the genotype miscalls and these are outlined next.

Examples of three different challenges for SNP genotyping with 34-plex are shown in Supplementary Fig. S1. First, SNPs rs10843344-rs239031 run to positions very close together, with the C peak of rs239031 often having a mobility shift that places it very close to the much higher C peak of rs10843344 (Supplementary Fig. S1.1). The same signal imbalance can be seen in the T peaks but the electrophoretic separation of these peaks remains more distinct. Examination of participant's SNaPshot profiles indicated some laboratories had missed the lower, shifted rs239031-C peak. Second, rs182549, rs881929 and rs3827760 have particularly low signal strengths (Supplementary Fig. S1.2) and the three SNPs show higher than average no-call rates. In the case of rs3827760, there is a very marked disparity in peak heights between the higher East Asian-informative G allele and the A allele (> 10:1 peak height ratio in the example shown), so this SNP requires particular care. Third, rs2304925 shows an artifactual G signal in the negative control very close to the G peak of rs5030240 (Supplementary Fig. S1.3). This peak is much higher than the T peak of rs2304925 when it is a true allelic extension product but much lower when artifactual. All participants ran a negative control and most recognized the extra G signal running close to the G peak of rs5030240, although as this is a tri-allelic SNP, when a homozygous A or C allele is present

the genotypes can be mistyped as an AG or CG in the absence of the stronger G peak with which to compare the artifact signal.

3.2. Genotyping performance of the Indel assay

All 19 participants successfully completed the genotyping of the samples with Indels. Supplementary Table S1 shows that almost half of the laboratories chose to dilute the PCR products 1:5-1:20 prior to CE detection to control signal pull up. Supplementary Figs. S1.4-5 show two examples of minor challenges with genotyping of Indels, consisting of the occurrence of dye blobs (broad non-specific peaks around allele peak positions), identifiable in the negative control, plus signal pull–ups that can occur when the Indel PCR products are not sufficiently diluted. However, there was no evidence that these two profile phenomena interfered with the genotyping performance of the Indel tests in any of the 19 laboratories. In fact, the genotyping completeness and concordance were very high when considering that most participants were running the test for the first time and required reading 46 different peak sets in each electropherogram.

Fig. 3B summarizes the Indel genotyping performance and shows participants achieved a very high overall genotyping completeness and concordance rate of 99.8%. Fourteen participants did not have miscalls or no-calls in any test sample profiles. A slight degree of clustering of genotyping miscalls and no-calls is discernible in Fig 3B; for example, laboratory #20 chose not to call 3/5 rs2307922 genotypes, and laboratories #1 and #7 mistyped more than one Indel. It is notable that all 19 participants successfully identified the rare third allele of rs25584 present in test sample C.

3.3. Inference of ancestry

All participants identified F as the mixed DNA sample and made Bayes analysis to infer the ancestry of samples A-E using *Snipper*. The majority, but not all, also made comparisons of the genotypes from A-E with the *Snipper* PCA module using the supplied reference population data. This section summarizes results for all laboratories using both statistical approaches to illustrate that the SNP and Indel data has a degree of ancestry-informativeness redundancy, i.e. the Bayes LRs or PCA positions of samples A-E are very similar despite some genotype miscalls or missing data. Therefore, the ancestry inferences made by participants were correct in all cases apart from those of laboratory #17 that made incorrect ancestry inferences for two samples and had PCA positions markedly displaced from the others in most cases.

Fig. 4A summarizes SNP profile quality (bar-charts, left-hand scale); Bayes LRs (points superimposed on bars, right-hand scale); and PCA positions for the SNaPshot assay data of 14 participants, analyzing samples A-E. Bayes LRs and PCAs from SNP data alone compare African, European and East Asian ancestries; consequently C and D give lower LRs and edge-of-cluster PCA positions that suggest East Asian ancestry despite these being Oceanian and American in origin. For 13/14 laboratories, samples A, B and E give mid-cluster PCA positions and high LRs that varied by four orders of magnitude between 1E+14 to 1E+18 correctly assigning A as East Asian and B as European, and 1E+22 to 1E+26 correctly assigning E as African. The LR values obtained by coordinating laboratory USC for SNP and Indel data are outlined in Table 1 (the 80-marker LRs for all samples are given separately in Fig. 2). Table 1 indicates sample C gave a high LR for Oceanian ancestry with just Indel data used in a 5-group comparison.

Fig. 4B summarizes Indel profile quality, Bayes LRs and PCA positions for a four group comparison using the Indel data of all participants. A sixth PCA plot, bottom right, shows the combined 80-marker analysis

for Oceanian sample C. Apart from African sample E, Indel data gives lower LRs than SNPs and the LRs for samples A and E are from different population likelihoods (bold values in Table 1). The improved genotyping consistency of Indels amongst participants is reflected in more uniform sets of Bayes LRs and PCA positions that mainly overlay each other (i.e. seen as single points on plots). For the two laboratories with three Indel miscalls, an effect is seen in the Bayes LRs for American sample D and African sample E, with some PCA displacement, indicating that even with just two markers miscalled it can sometimes affect the statistical inference made from other correctly called genotypes (~97% of the data). The Oceanian sample C was correctly identified by 18 participants, with many using both Indel and combined data to make the inference.

Therefore, 18 of 19 laboratories were able to successfully assign ancestry to five samples of undisclosed geographic origin, obtaining unequivocal Bayes LRs and, in most, cases participants constructed PCA plots providing supplementary analyses with good matches to the Bayes results.

3.4. Mixture detection and analysis of participant's Indel peak height data

Although the exercise was not a fully blinded test (i.e. where the presence of a mixed sample is not disclosed), all participants were able to identify sample F as the mixture from the observation of imbalanced signals in the heterozygote peak pairs of the Indel profile. Therefore, despite a lack of familiarity with Indel peak patterns in most laboratories, there was sufficient contrast between the mixed sample F and the unmixed A-E DNAs for the mixture to be discernible by all participants. In addition, 7 of 14 laboratories reported an ACG triple-peak pattern in the tri-allelic SNP rs5030240, one reported an AC with possible G, one a GG result and the other five gave no-calls. A typical sample F peak pattern for rs5030240 is shown in Supplementary Fig. S1.6.

The detection of peak height imbalances that can indicate mixed DNA has been stated to be an advantage of direct PCR-to-CE Indel genotyping compared to SNaPshot tests [15,17,19], however such patterns have not been properly assessed across a range of CE detectors. For this reason, we decided to ask participants to provide their heterozygote peak height data and then compiled the variation in peak height ratios (PHRs, highest/lowest peaks) recorded in the five unmixed and single mixed DNAs from the range of CE regimes used. Furthermore, when analyzing binary markers the number of heterozygotes observed in mixtures is invariably higher than normal unmixed samples. Although PHR values were distinct between A-E and F, three factors complicated the straightforward statistical comparison of patterns of heterozygosity observed amongst the test samples. First, there was variation in the number of heterozygotes recorded in sample F. Specifically, laboratory #1 identified 18 heterozygotes; #15: 17; #18: 21; and #20: 17, compared to an average number of heterozygotes identified by the other fifteen laboratories of 27. Second, the lower number of identified heterozygotes for F in some participant's data affects the minimum-maximum and average PHR values, particularly when the PHR is extreme and a very low peak is discounted when reading the profile. Four example peak pairs that were recorded as single allele genotypes by one participant but as heterozygotes by the others, are shown in Fig. 5A. Third, due to the contrasting frequencies of most of the 46 Indels between population groups, sample A showed lower numbers of heterozygotes and sample B higher numbers than those seen in C-E.

The numbers of heterozygotes and PHR values are plotted in Fig. 5B. This chart shows data from 15/19 laboratories (excluding #1, #15, #18 and #20). The same chart with all 19 participant's data is shown in Supplementary File S4.A. The dark grey bars mark the data from 3500 detectors and indicates that no difference in peak height ratios are discernible in comparison with 31xx CE data.

Statistical assessment of the number of heterozygotes in A-E vs. F was made with a unilateral 2-sample test for equality-of-proportions (with continuity correction). The resulting grid of *p*-values for pairwise comparisons across all 19 laboratories is shown in Supplementary File S4.B, along with the Fig 5B chart re-plotted for full data from all laboratories (Supplementary File S4.A). It can be seen from the Supplementary File S4.A chart that the numbers of sample F heterozygotes recorded by laboratories #1, #15 and #20 is lower than the average number in unmixed sample B. Inclusion of this data has a direct effect on the distribution of significant *p*-values obtained from pairwise comparisons. Laboratories #1, #15 and #20 sample F heterozygote numbers are significantly different to those of most of the other laboratories, but not different to heterozygote numbers in unmixed samples B-E, while #18 data for sample F is not significantly different to samples B and C. The high number of heterozygotes in sample B is reflected in significant differences only found for comparisons to those of laboratories #8, #13, #14, and #5, who recorded 29 or more heterozygotes in their sample F profiles. Therefore, we opted to remove #1, #15, #18 and #20 data from the statistical assessment of PHR differences between A-E and F.

The average PHRs shown in Fig. 5B indicate a quite distinct contrast between samples A-E and F, with values of 1.15 compared to 3.14 respectively, which suggests a ratio of 1:2.73 that approximates the actual 1:3 contributor ratio well. Although the PHR values give a clearly discernible difference between mixed and unmixed samples, we completed a formal statistical test of this difference. An ANOVA test is a standard approach for assessing continuous values such as PHR measurements, but a Shapiro-Wilks test indicated some of the data was not normally distributed (data not shown). Therefore, a Kruskal-Wallis rank sum test was applied and the grid of pairwise *p*-values comparing the average PHRs of A-E with individual PHRs of F is shown in Supplementary File S4.C. The results are completely consistent: the pairwise comparisons of mixed vs. unmixed PHRs give significant *p*-values in every case and none were detected for comparisons within each sample set.

In summary, despite the need to adjust statistical comparisons by removing 4 of 19 participant's data due to under-reported heterozygote peak pairs, the other laboratories provided a ratio of average peak heights close to 1:3. This ratio is consistent with the mixture that was constructed for the exercise and is statistically significant for all signal strength comparisons made.

3.5. Additional Next Generation Sequencing experiments applied to test DNAs by two laboratories

Two laboratories decided to use their remaining PCR primers to genotype one or both marker sets with different Next Generation Sequencing (NGS) systems, as outlined in Supplementary File S5. One assessed 34-plex SNP typing using an unmodified PCR followed by library preparation and massively parallel sequencing with the Illumina MiSeq system. The other assessed 34-plex SNPs and Indel genotyping in the same way (unmodified PCR in each case) with the Thermo Fisher Scientific-Life Technologies (TFS-LT) Ion PGM™ system.

The 34-plex SNP sequence analyses were successful to a very large degree, as all genotypes were identified and almost fully concordant with each laboratory's SNaPshot data. Sample F was observed to be distinct in a major proportion of its allele-pair sequence ratios (defined as the second allele exceeding 10% of sequence reads), compared to A-E. Supplementary File S5 indicates there were only 5/14 sequence ratios of 1.5 or less (i.e. in the range: 0.4:0.6-0.5:0.5) in the Ion PGM™ data and 3/17 in the MiSeq data. This equates to 64% and 82% of sequence ratios exceeded those of most normal DNA heterozygotes seen in Ion PGM™ and MiSeq respectively, giving unequivocal signals of a mixture in F. Both systems also detected displaced sequence ratios in each of the two tri-allelic SNPs.

The Indel analysis with NGS gave three discordant genotypes in samples B and C, plus an average 8.7% no-calls (coverage too low) and 2.9% missing data (undetected sequence), although not all samples gave the same non-detection rates. Overall, 84% of the NGS genotypes matched the CE calls. However, the alignment of sequences that contain short insertions and deletions is particularly challenging in NGS sequence analysis and it was not possible to be sure how many miscalls or no-calls were due to misalignment issues. Supplementary File S5 shows assessments of Indel sequence ratios for sample F compared to A-E. Given that sequence coverage was low in some loci and this is the first NGS experiment with this type of forensic marker, results need cautious interpretation. However, patterns suggest a degree of displacement in F away from the perfect sequence balance midline (0.5:0.5) compared to many of the heterozygote sequence ratios detected in A-E.

4. Discussion

As forensic NGS analysis gains greater traction, it is the right moment for the forensic community to use inter-laboratory exercises to assess the binary marker sets that will start to add complementary genetic data to conventional STR polymorphisms. Ancestry inference is seen as a key part of the enhanced characterization of forensic DNA that NGS will allow. Therefore, it is important to evaluate the robustness of existing CE-based ancestry-informative SNP and Indel multiplexes in terms of how easily they can be adopted in laboratories not previously experienced with binary marker genotyping. The statistical analysis of the genotype data obtained from AIMs also needs to be easy to use and interpret by forensic laboratories. The most straightforward approach for inferring ancestry uses Bayesian LR comparisons between the two geographic origin hypotheses with the highest likelihoods. Lastly, binary variation has a reduced capacity to detect mixtures since homozygotes in combination can look like heterozygotes and only a few non-binary SNPs or Indels currently offer the chance to observe more than two alleles. Therefore, the exercise findings for genotyping reliability, ease-of-use of the recommended ancestry inference tools and ability to detect mixed DNA are all relevant to the progress towards adoption of AIMs in forensic analysis.

The principal finding of this exercise was that each of the participants readily established the AIM-Indel 46-plex test in their laboratory. All participants achieved good quality profiles that reached the high level of genotyping concordance of 99.8% and then efficiently detected mixed sample F. In contrast, SNaPshot typing was both more challenging and for many participants less reliable, despite most laboratories having successfully genotyped six SNPs for the preceding *IrisPlex* EDNAP exercise [21]. Miscalled genotypes with SNaPshot produced an overall genotyping concordance rate of 97.3% when a single participant's results were excluded (13/14 laboratories).

We have no explanation for the very high number of SNP miscalls from this one laboratory but it resulted in their statistical analyses producing the only incorrect ancestry inferences for two test DNAs and one uninformative LR of 1.2. All other participants produced correct ancestry predictions from the Bayes LRs calculated in *Snipper* and, for those that created PCA plots, obtained cluster patterns and profile positions that corresponded to these LRs. Therefore, from the review of exercise reports returned from 19 laboratories, we can recommend the use of both of these statistical approaches to ancestry inference, as these proved easy to use and allowed correct ancestry assignments of samples with undisclosed geographic origin in 92 of 95 cases.

Mixture detection achieved from Indel peak patterns was particularly successful, with sample F giving a clear signal of mixed DNA for all participants. Our analysis of peak height ratios made after the exercise finished, gave a good approximation of the actual mixture component ratio, averaging 1.15 and 3.14 for PHRs in unmixed samples and the mixture respectively. The much higher number of heterozygotes in F could mainly be due to the different ancestries of the mixture contributors. Nevertheless, recording a higher number of heterozygotes than in normal DNA samples and observing PHRs markedly above ~1.2 gives a simple and easily adopted system to detect mixtures with Indels. SNaPshot does not offer the same direct relationship between peak heights and input DNA so there is a risk that simple two-person mixtures mimic the patterns seen in individuals with co-ancestry due to population admixture, as revealed by the PCA plot of sample F in Fig. 2 (plot A). Obviously, single sample experiments are not fully indicative of how well Indels will perform with a range of forensic samples, mixture ratios or component ancestry combinations, but the fact that most participants were running Indels for the first time and all detected

401 the mixture indicates sensitivity to mixed DNA with this assay.

402 Although the NGS findings from two participants are a set of parallel genotyping experiments using
403 exercise materials that were not part of the study plan, results are included in this report to highlight the
404 enhanced sensitivity to mixtures obtained for SNP analysis with NGS. It is also interesting to note that
405 existing optimized forensic multiplexes work very well in NGS without the need for any modification,
406 confirming the results of a recent study that found the 34-plex PCR primers, amongst four other forensic
407 SNP multiplexes, provide good quality output with the Ion PGM™ system [24]. In addition, the relative
408 success of the initial Indel genotyping experiments with NGS indicate dye-labeled PCR primers do not
409 interfere with library preparation and subsequent sequencing chemistry of the Ion PGM™. This suggests
410 existing forensic CE multiplexes for a range of markers, including STR kits, could be used to prepare target
411 DNA for experimental NGS sequencing runs.

412 Until NGS systems that incorporate AIMs are widely adopted for forensic use, the results from this EDNAP
413 inter-laboratory exercise indicate the PCR-to-CE Indel test is by far the best current option for forensic
414 ancestry analysis. The Indel multiplex provides a simple, reliable and informative test from a
415 comparatively large marker set that is analyzed using validated CE regimes. Detection of simple two-
416 component mixed DNA from scrutiny of Indel peak patterns was a task accomplished by all exercise
417 participants and gives Indel genotyping a key additional advantage over SNP-based ancestry tests.

418 **Acknowledgements**
419
420 This work was partially supported through a grant to CS (SFRH/BD/75627/2010) awarded by the
421 Portuguese Foundation for Science and Technology (FCT) and co-financed by the European Social Fund
422 (Human Potential Thematic Operational Programme). The organizing laboratory of the Forensic Genetics
423 Unit, University of Santiago de Compostela, would like to thank Fabio Oldoni for his assistance in piloting
424 the exercise, in partial fulfillment of an MSc in Forensic Science, Department of Forensic Science and Drug
425 Monitoring, King's College, London. The two anonymous donors for the mixture are also thanked for their
426 willingness to provide samples. Zurich Institute of Forensic Medicine thanks Corinne Moser for technical
427 assistance. The pilot studies, Indel primers and test DNAs used in the exercise were funded by the
428 European Forensic Genetics Network of Excellence (EUROFORGEN-NoE, Grant Agreement No. 285487).

429 **References**
430
431 [1] C. Phillips, L. Prieto, M. Fondevila, A. Salas, A. Gomez-Tato, J.A. Alvarez-Dios, A. Alonso, A. Blanco-
432 Verea, M. Brión, M. Montesino, Á. Carracedo, M.V. Lareu, Ancestry analysis in the 11-M Madrid bomb
433 attack investigation. PLoS One 4 (2009) e6583.

434 [2] S. Walsh, F. Liu, A. Wollstein, L. Kovatsi, A. Ralf, A. Kosiniak-Kamysz, W. Branicki, M. Kayser, The
435 HIrisPlex system for simultaneous prediction of hair and eye colour from DNA, Forensic Sci. Int. Genet. 7
436 (2013) 98–115.

437 [3] M. Kayser, P. de Knijff, Improving human forensics through advances in genetics, genomics and
438 molecular biology, Nat. Rev. Genet. 12 (2011) 179–192.

439 [4] M. Kayser, P.M. Schneider, DNA-based prediction of human externally visible characteristics in
440 forensics: motivations, scientific challenges, and ethical considerations, Forensic Sci. Int. Genet. 3 (2011)
441 154–161.

442 [5] C. Phillips, L. Fernandez-Formoso, M. Garciañas, L. Porras, T. Tvedebrink, J. Amigo, M. Fondevila, A.
443 Gomez-Tato, J. Alvarez-Dios, A. Freire-Aradas, A. Gomez-Carballa A. Mosquera-Miguel, Á. Carracedo, M.V.
444 Lareu, Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using
445 the CEPH human genome diversity panel, Forensic Sci. Int. Genet. 5 (2011) 155–169.

446 [6] C. Phillips, M. Gelabert-Besada, L. Fernandez-Formoso, M. García-Magariños, C. Santos, M.Fondevila,
447 D. Ballard, D. Syndercombe Court, Á. Carracedo, M.V. Lareu, "New turns from old STaRs": Enhancing the
448 capabilities of forensic short tandem repeat analysis, Electrophoresis 35 (2014) 3173-3187.

449 [7] T.E. King, E.J. Parkin, G. Swinfield, F. Cruciani, R. Scozzari, A. Rosa, S.K. Lim, Y. Xue, C. Tyler-Smith, M.A.
450 Jobling, Africans in Yorkshire? The deepest-rooting clade of the Y phylogeny within an English genealogy,
451 Eur. J. Hum. Genet. 15 (2007) 288-293.

452 [8] O. Lao, P.M. Vallone, M.D. Coble, T.M. Diegoli, M. van Oven, K.J. van der Gaag, J. Pijpe, P. de Knijff, M.
453 Kayser, Evaluating self-declared ancestry of U.S. Americans with autosomal, Y-chromosomal and
454 mitochondrial DNA, Hum. Mutat. 31 (2010) E1875–E1893.

455 [9] D. Corach, O. Lao, C. Bobillo, K. van Der Gaag, S. Zuniga, M. Vermeulen, K. van Duijn, M. Goedbloed,
456 P.M. Vallone, W. Parson, P. de Knijff, M. Kayser, Inferring continental ancestry of Argentineans from
457 autosomal, Y-chromosomal and mitochondrial DNA, Ann. Hum. Genet. 74 (2010) 65–76.

458 [10] C. Bouakaze, C. Keyser, E. Crubézy, D. Montagnon, B. Ludes, Pigment phenotype and biogeographical
459 ancestry from ancient skeletal remains: inferences from multiplexed autosomal SNP analysis, Int. J. Leg.
460 Med. 123 (2009) 315-325.

461 [11] M. Fondevila, C. Phillips, C. Santos, A. Freire Aradas, P.M. Vallone, J.M. Butler, M.V. Lareu, Á.
462 Carracedo, Revision of the SNP*for*ID 34-plex forensic ancestry test: assay enhancements, standard
463 reference sample genotypes and extended population studies, Forensic Sci. Int. Genet. 7 (2013) 63-74.

464 [12] C. Phillips, A. Freire Aradas, A.K. Kriegel, M. Fondevila, O. Bulbul, C. Santos, F. Serrulla Rech, M.D.

Perez Carceles, Á. Carracedo, P.M. Schneider, M.V. Lareu, Eurasiaplex: A forensic SNP assay for differentiating European and South Asian ancestries, Forensic Sci. Int. Genet. 7 (2013) 359–366.

[13] M. Fondevila, C. Phillips, N. Naveran, L. Fernandez, M. Cerezo, A. Salas, Á. Carracedo, M.V. Lareu, Case report: identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur, Forensic Sci. Int. Genet. 2 (2008) 212–218.

[14] C. Børsting, E Rockenbauer, N. Morling, Validation of a single nucleotide polymorphism (SNP) typing assay with 49 SNPs for forensic genetic testing in a laboratory accredited according to the ISO 17025 standard, Forensic Sci. Int. Genet. 4 (2009) 342–42.

[15] R. Pereira, C. Phillips, C. Alves, A. Amorim, Á. Carracedo, L. Gusmão, A new multiplex for human identification using insertion/deletion polymorphisms, Electrophoresis 30 (2009) 3682–3690.

[16] S.L. Friis, C. Børsting, E. Rockenbauer, L. Poulsen, S.F. Fredslund, C. Tomas, N. Morling, Typing of 30 insertion/deletions in Danes using the first commercial indel kit—Mentype® DIPplex, Forensic Sci. Int. Genet. 6 (2012) e72-e74.

[17] M. Fondevila, C. Phillips, C. Santos, R. Pereira, L. Gusmão, Á. Carracedo, J.M. Butler, M.V. Lareu, P.M. Vallone, Forensic performance of two insertion-deletion marker assays, Int. J. Legal Med. 126 (2012) 725-737.

[18] N.P. Santos, E.M. Ribeiro-Rodrigues, A.K. Ribeiro-dos-Santos, R. Pereira, L. Gusmão, A. Amorim, J.F. Guerreiro, M.A. Zago, C. Matte, M.H. Hutz, S.E. Santos, Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INDEL) ancestry- informative marker (AIM) panel. Hum. Mutat. 31 (2010) 184–190.

[19] R. Pereira, C. Phillips, N. Pinto, C. Santos, C.E.B. Santos, A. Amorim, A. Carracedo, L. Gusmão, Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing, PLoS One 7 (2012) e29684.

[20] D. Zaumsegel, M.A. Rothschild, P.M. Schneider, A 21 marker insertion deletion polymorphism panel to study bio-geographical ancestry, Forensic Sci. Int. Genet. 7 (2013) 305-312.

[21] L. Chaitanya, S. Walsh, J.D. Andersen, R. Ansell, K. Ballantyne, D. Ballard, R. Banemann, C.M. Bauer, A.M. Bento, F. Brisighelli, et al., Collaborative EDNAP exercise on the IrisPlex system for DNA-based prediction of human eye colour, Forensic Sci. Int. Genet. 11 (2014) 241-251.

[22] J. Novembre, M. Stephens, Interpreting principal component analyses of spatial population genetic variation, Nat. Genet. 40 (2008) 646-649.

[23] D. Reich, A. Price, N. Patterson, Principal component analysis of genetic data, Nat. Genet. 40 (2008) 491–492.

[24] R. Daniel, C. Santos, C. Phillips, M. Fondevila, R.A. van Oorschot, Á. Carracedo, M.V. Lareu, D. McNevin, A SNaPshot of next generation sequencing for forensic SNP analysis, Forensic Sci. Int. Genet. 14 (2015) 50–60.

Table 1. Lowest LR values produced from *Snipper* Bayes analysis of the full SNP and Indel profiles of samples A-E and 9947A with their ancestry inferences. Participant LR values for the same samples are plotted in Figs. 4A/B. Bold values for A and E highlight different population ratios giving the lowest LRs when SNP, 3-group or Indel, 4-group comparisons are made. With Indel, 4-group comparisons the second lowest LRs for samples A and E are based on the same population ratios as the lowest LRs for SNP, 3-group comparisons. Sample C is correctly inferred to be Oceanian with Indel data alone but most participants reported the LR from 80 marker data.

Inference:	**34-plex SNPs, 3-group**	
European	9947A is 2,118,840,589,047,061,020,672 times more likely EUROPEAN than E ASIAN	
East Asian	**A is 361,148,635,069,545,024 times more likely E ASIAN than EUROPEAN**	
European	B is 64,191,487,284,485,608 times more likely EUROPEAN than E ASIAN	
East Asian	C is 13,115,706 times more likely E ASIAN than AFRICAN	
East Asian	D is 248,539,593,557 times more likely E ASIAN than EUROPEAN	
African	**E is 556,454,701,312,037,054,117,314,560 times more likely AFRICAN than E ASIAN**	
	46-plex Indels, 4-group	**46-plex, 4-group (second lowest LR)**
European	9947A is 1,937,432,967,198 times more likely EUROPEAN than E ASIAN	
East Asian	**A is 6,993,957 times more likely E ASIAN than AMERICAN**	A is 37,290,377,821,078,192,128 times more likely E ASIAN than EUROPEAN
European	B is 143,659,679,122 times more likely EUROPEAN than E ASIAN	
LR too low	C is 131 times more likely E ASIAN than EUROPEAN	
American	D is 944,698,134 times more likely AMERICAN than E ASIAN	
African	**E is 3,229,841,442,838,053,650,432 times more likely AFRICAN than EUROPEAN**	E is 5,715,694,248,335,998,122,459,136 times more likely AFRICAN than E ASIAN
	46-plex, 5-group	**80 Markers, 5-group**
Oceanian	C is 24,880,402 times more likely OCEANIAN than E ASIAN	C is 153,747,536,542,653 times more likely OCEANIAN than E ASIAN

Figure legends

Fig. 1. Electropherograms from the Indel test (upper panel) and the 34-plex SNP test for the 9947A control DNA. Peak positions are labeled with the internal codes used for each marker (internal code-rs-number lists are provided in Fig. 3A; Supplementary Files S3; *Snipper* and in [11]).

Fig. 2. Ancestry analysis of exercise test samples. 80-marker genotypes were analyzed and HGDP-CEPH training set data was as supplied to participants (Supplementary File S3.2). Top plot shows ranked *Snipper* Bayes analysis LRs from training set cross validation or test profile analysis (black points). Grey points in East Asians/Oceanians indicate LRs below a threshold value of 1000 (the grey shaded log LR range around balanced odds line of LR=1). Red points indicate East Asian training set LRs that misclassified as Americans. Middle plots show *STRUCTURE* analysis aligned directly to the LR distributions above with separate plots for mixture components, left and test samples, right. Lower plots show 2D PCA analyses of test samples in 3-group or 5-group comparisons. Plot A shows a 3-group comparison of sample F, positioned mid-cluster between contributors M1 and M3. Plot B shows the full 5-group PCA of samples A-E plus 9947A. Plot C shows a restricted comparison of just East Asian, Oceanian and American data to obtain better differentiation of reference population clusters and A, C, D; all more closely distributed in plot B.

Fig. 3. (A) Genotyping performance of the 34-plex test arranged by SNP (rows) and by 14 participants (columns). Cells record miscalls on the left, and no-calls right. The bar plots on the right summarize total genotype completeness and concordance for each SNP and at the bottom, for each participant. SNPs are ordered by diminishing performance (i.e. decreasing concordance then completeness). Overall genotype concordance is given for 14 and 13 laboratories separately, excluding participant #17 with a very high number of SNP miscalls. (B) Genotyping performance grid for Indel test data from all 19 laboratories. Miscalls are shown as dark grey cells, no-calls light grey.

Fig. 4. (A) Participant's SNP-based Bayes LRs and PCA positions for three-group comparisons (AFR-EUR-E ASN) analyzing samples A-E. Genotype completeness and concordance rates are shown as bar charts (left-hand scales) and ancestry assignment LRs (i.e. lowest values) as overlaid points (right-hand scales). Laboratories with some displacement of a sample position from the main PCA cluster are individually labeled and incorrect positions/assignments from miscalled genotypes are shown in red. (B) Participant's Indel-based Bayes LRs and PCAs for 4-group comparisons (including Americans) analyzing A-E. The sixth plot, lower right, shows a 5-group PCA of sample C (adding Oceanian reference data) using 80-marker genotypes. Laboratories only reporting Indel data have Bayes LRs shown in green and one uninformative LR shown in blue.

Fig. 5. (A) Example Indel peak pairs for sample F discounted as heterozygotes by one participant. (B) Numbers of Indel heterozygotes (bars) and their peak height ratios (PHR: points) recorded by 15 participants. Unmixed samples A-E are average values from all data and sample F values are shown individually as different numbers of peak pairs were recognized as heterozygotes amongst laboratories.

Supplementary Files

Supplementary Figs. S1. Examples of genotyping challenges in 34-plex or Indel profiles.

Supplementary Table S1. Capillary electrophoresis (CE) details for participating laboratories. Grey bars for 34-plex denote five participants not completing SNP genotyping with SNaPshot.

Supplementary File S1. Laboratory protocol guide in the form of an Excel calculator for reaction setups provided to exercise participants.

Supplementary File S2. CE fragment mobility panels-and-bins files provided to exercise participants.

Supplementary Files S3. SNP and Indel genotypes used in the exercise as reference population data plus test DNA data established by USC. Worksheets are:
File S3.1 PCA input
File S3.2 The 5-group training set data for Bayes analysis
File S3.3 Cross-validation data plotted in Fig. 2.
Note that to use the files for *Snipper* analysis each must be moved to 'worksheet position 1'.
PCA: http://mathgene.usc.es/snipper/analysismultipleprofiles.html,
Bayes custom or fixed training set data: http://mathgene.usc.es/snipper/analysispopfile_new.html http://mathgene.usc.es/snipper/popchoosing5groups.html).

Supplementary File S4. Statistical analysis of participant's Indel heterozygote peak height ratio data.

Supplementary File S5. Details and results of NGS analyses of 34-plex and Indel markers made by two laboratories.

Figure 1

Click here to download high resolution image

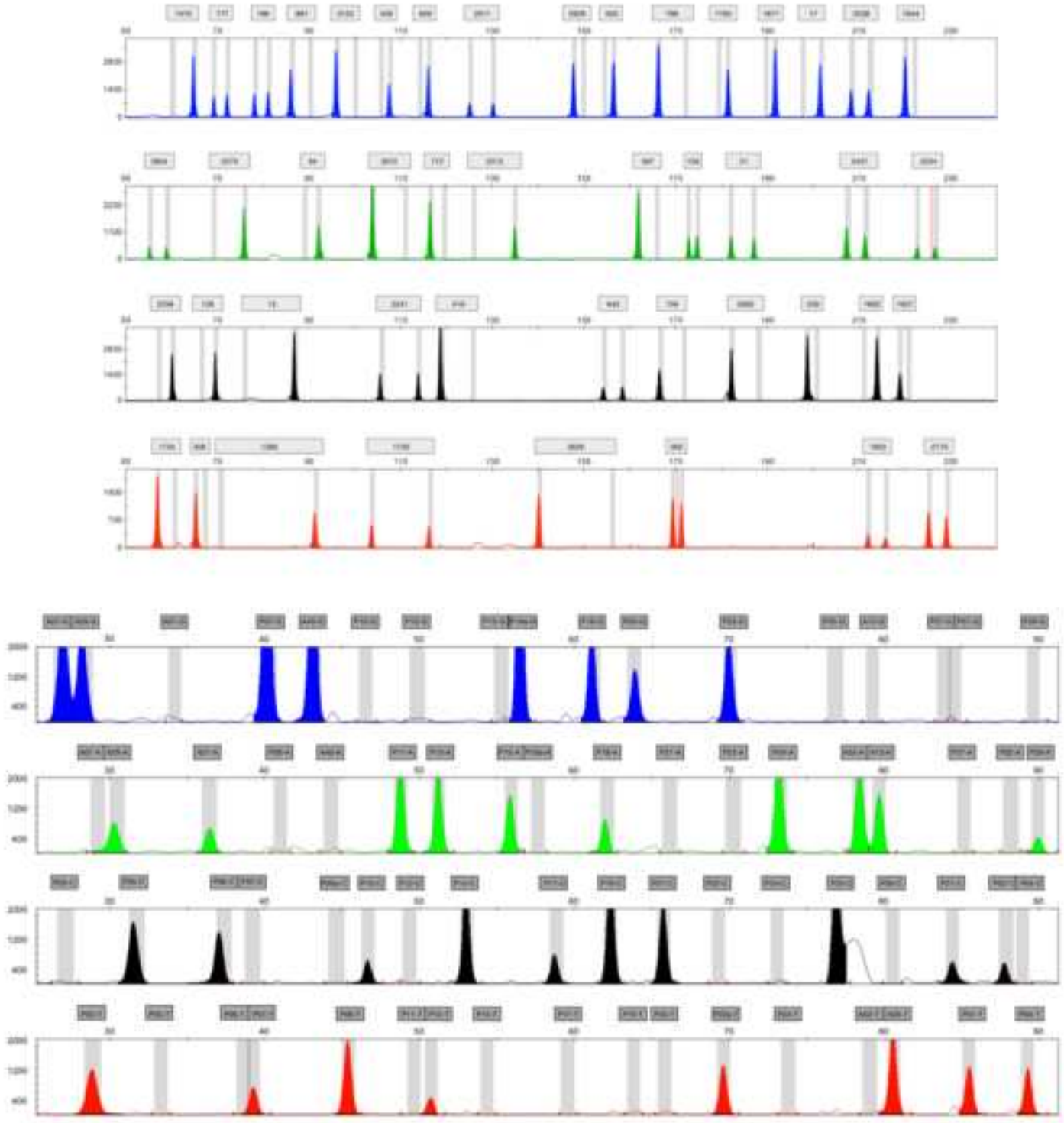


Figure 2

Click here to download high resolution image

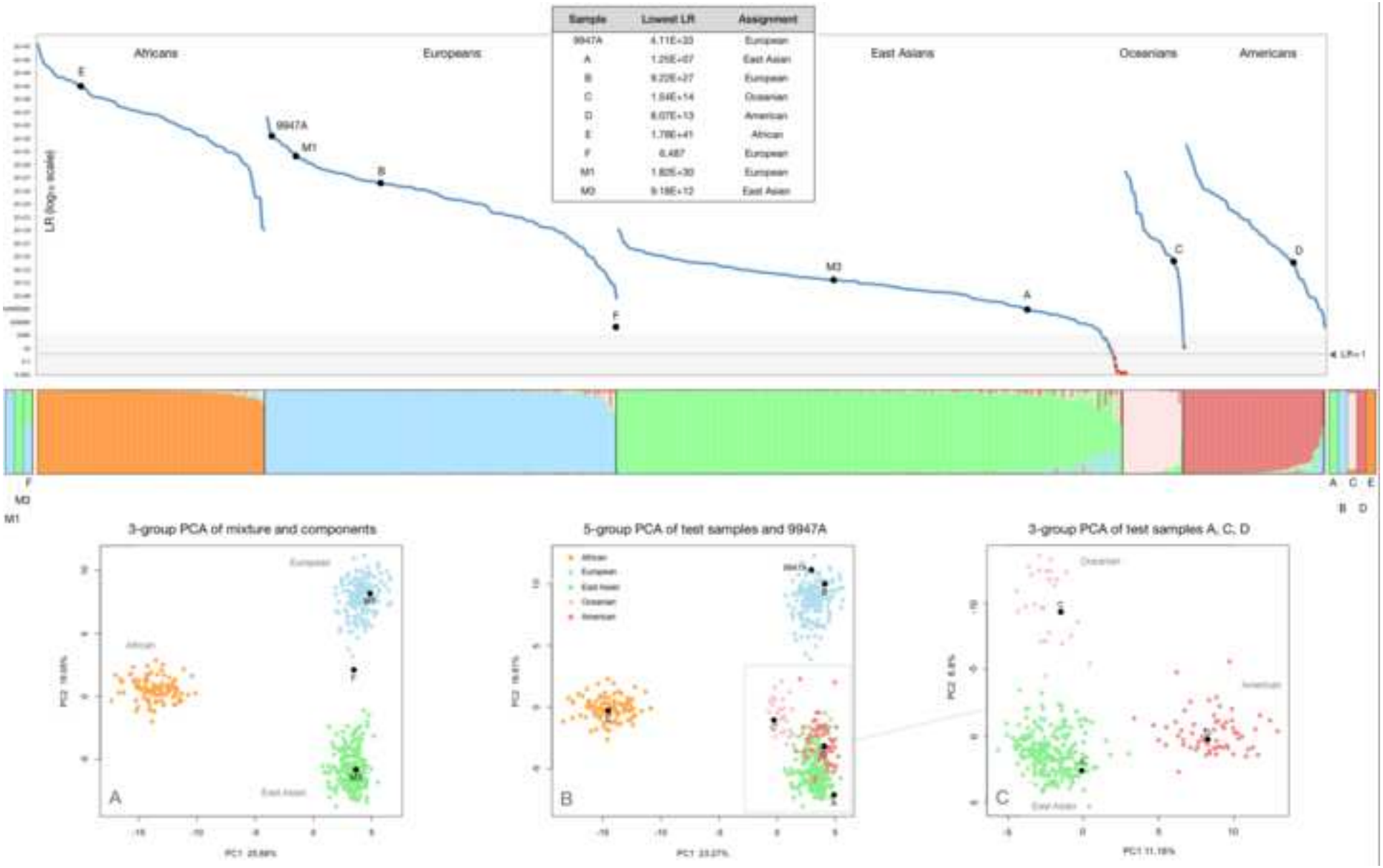


Figure 3A

Click here to download high resolution image

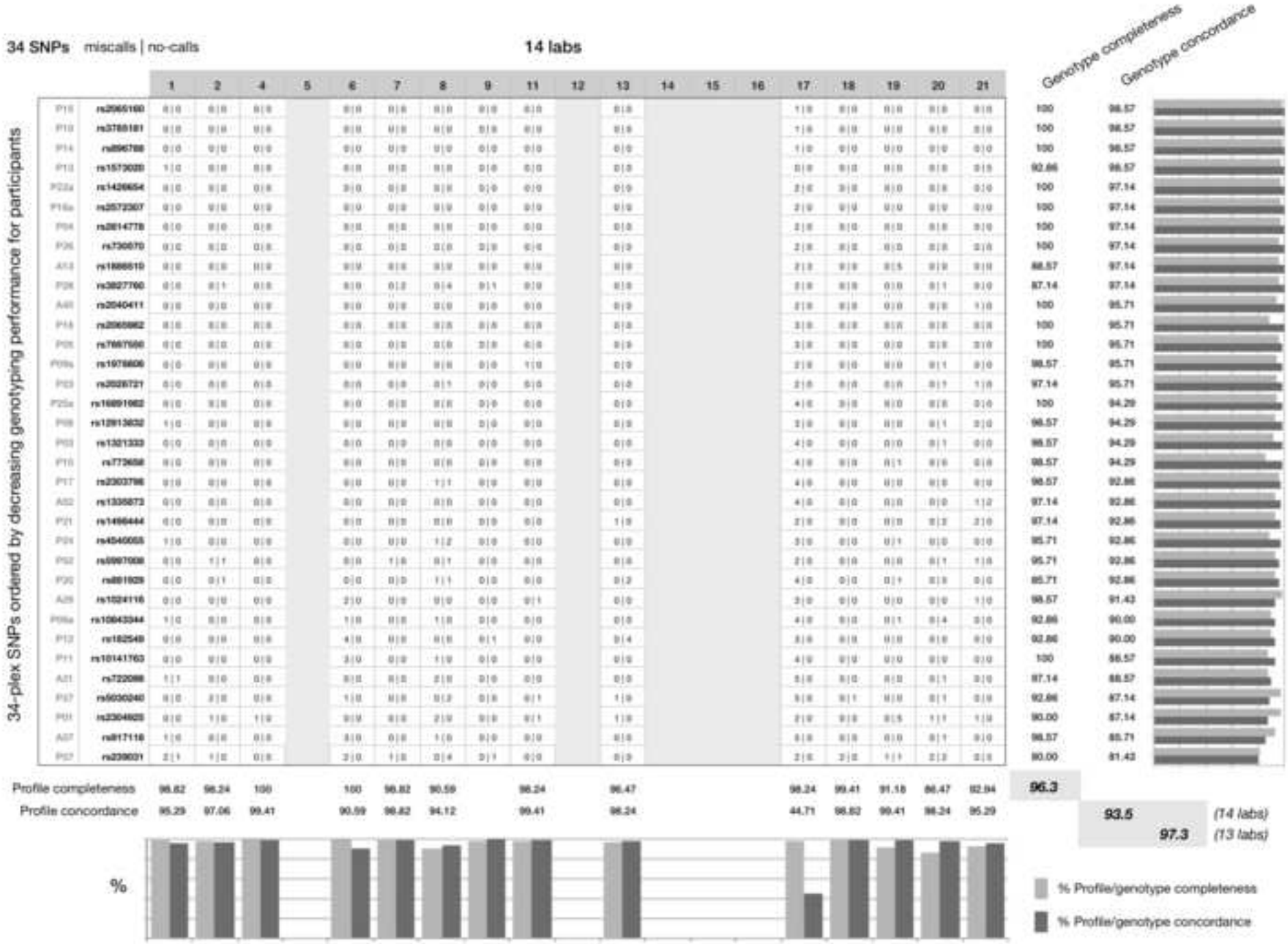


Figure 3B

Click here to download high resolution image

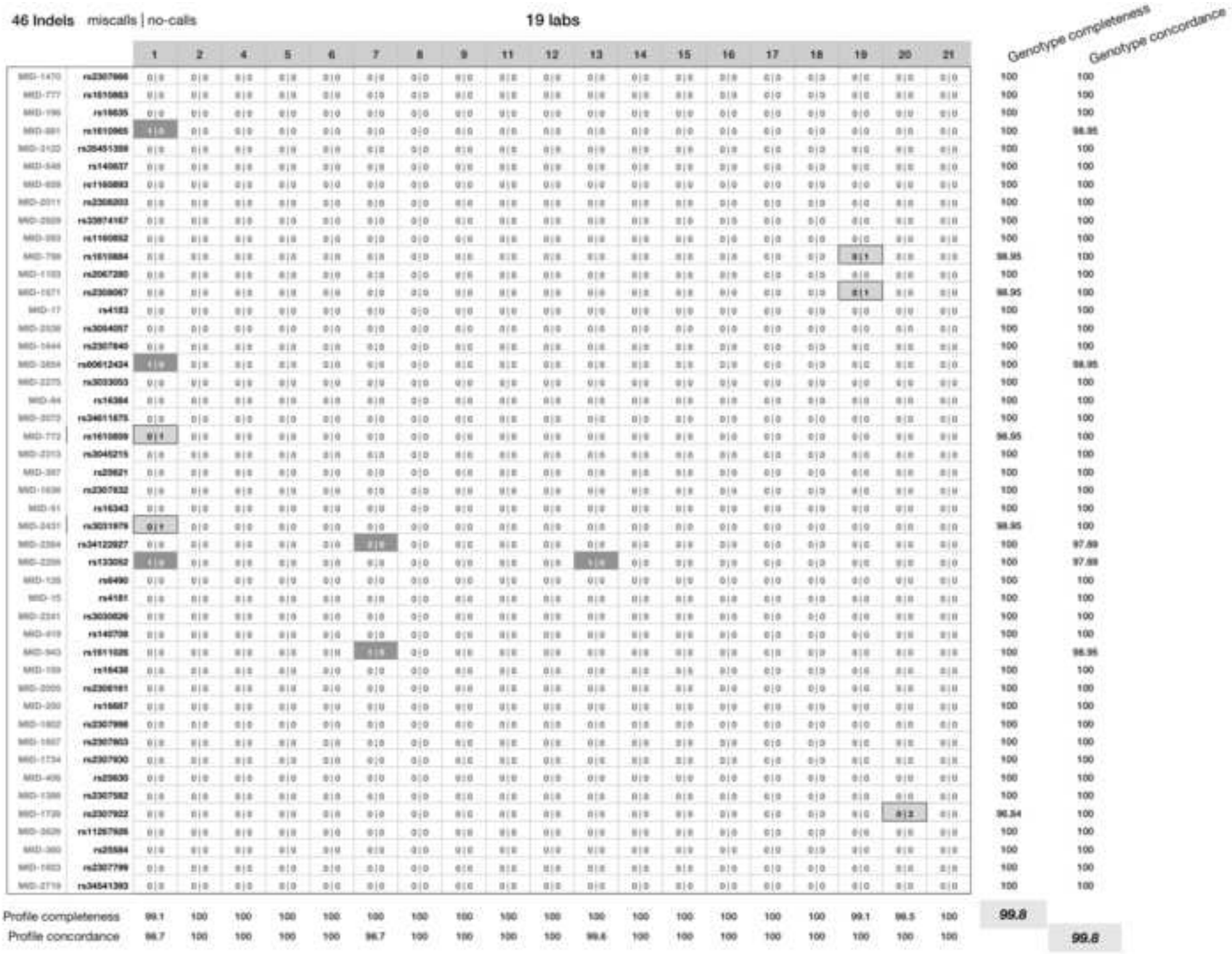


Figure 4A

Click here to download high resolution image

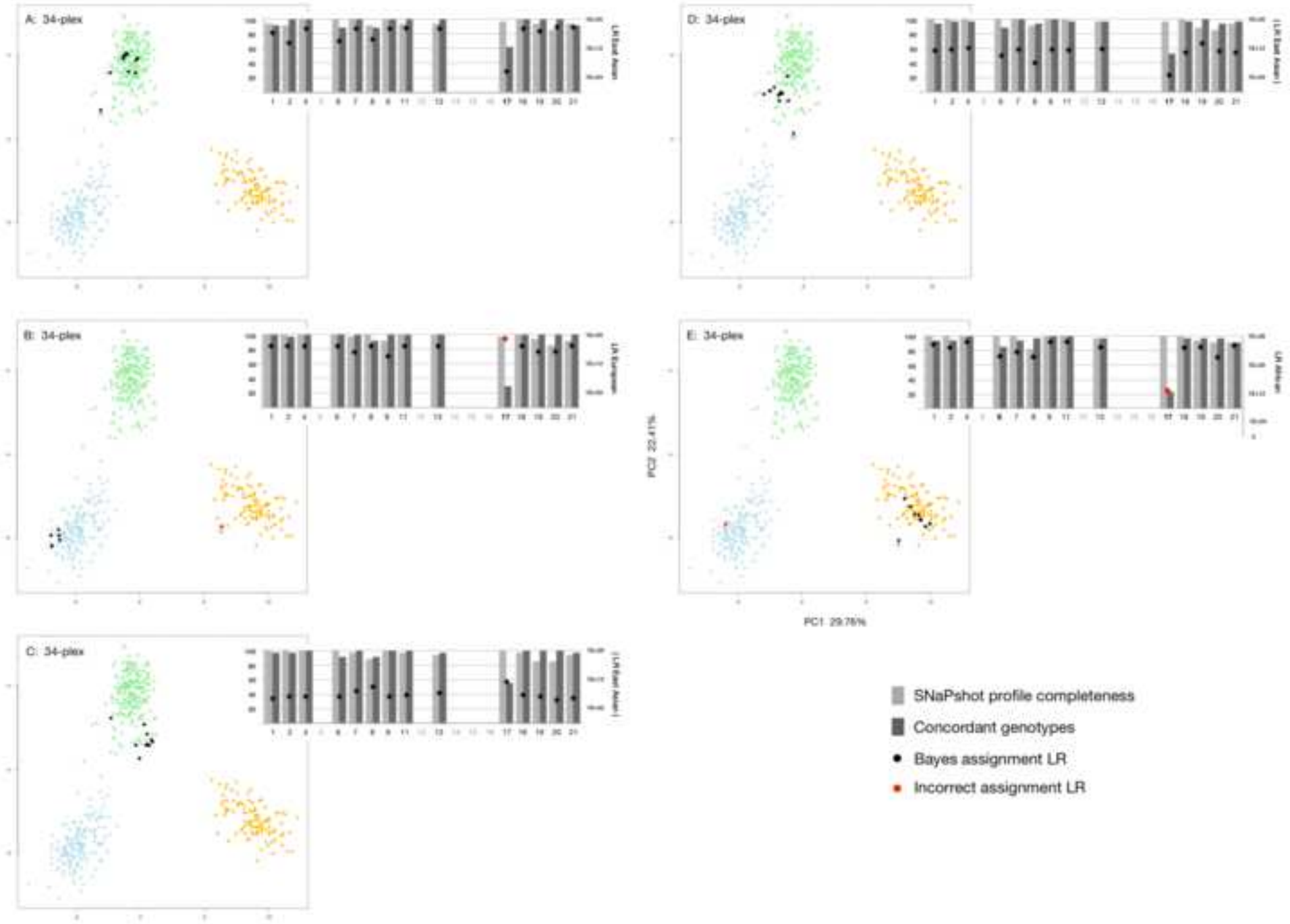


Figure 4B

Click here to download high resolution image

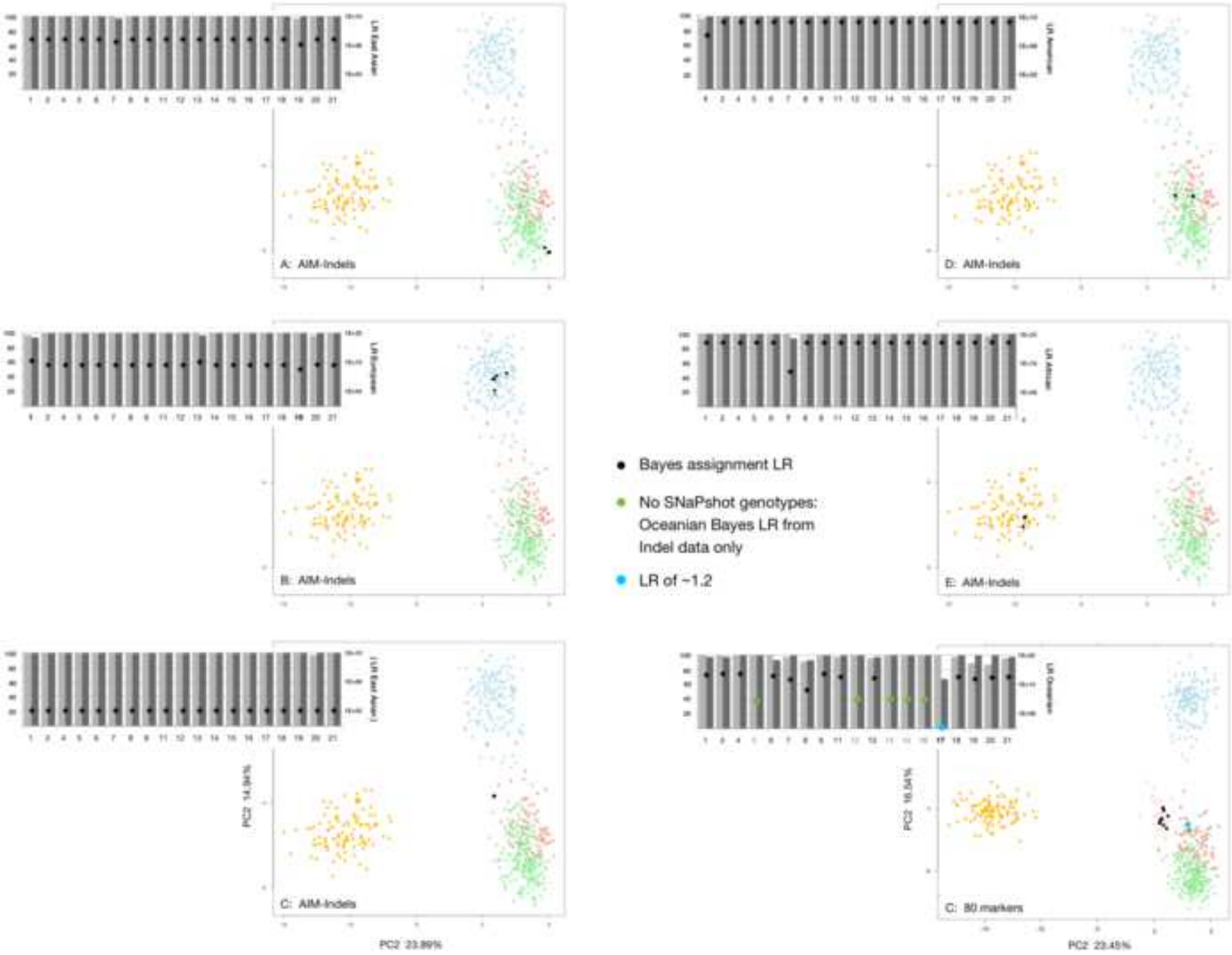
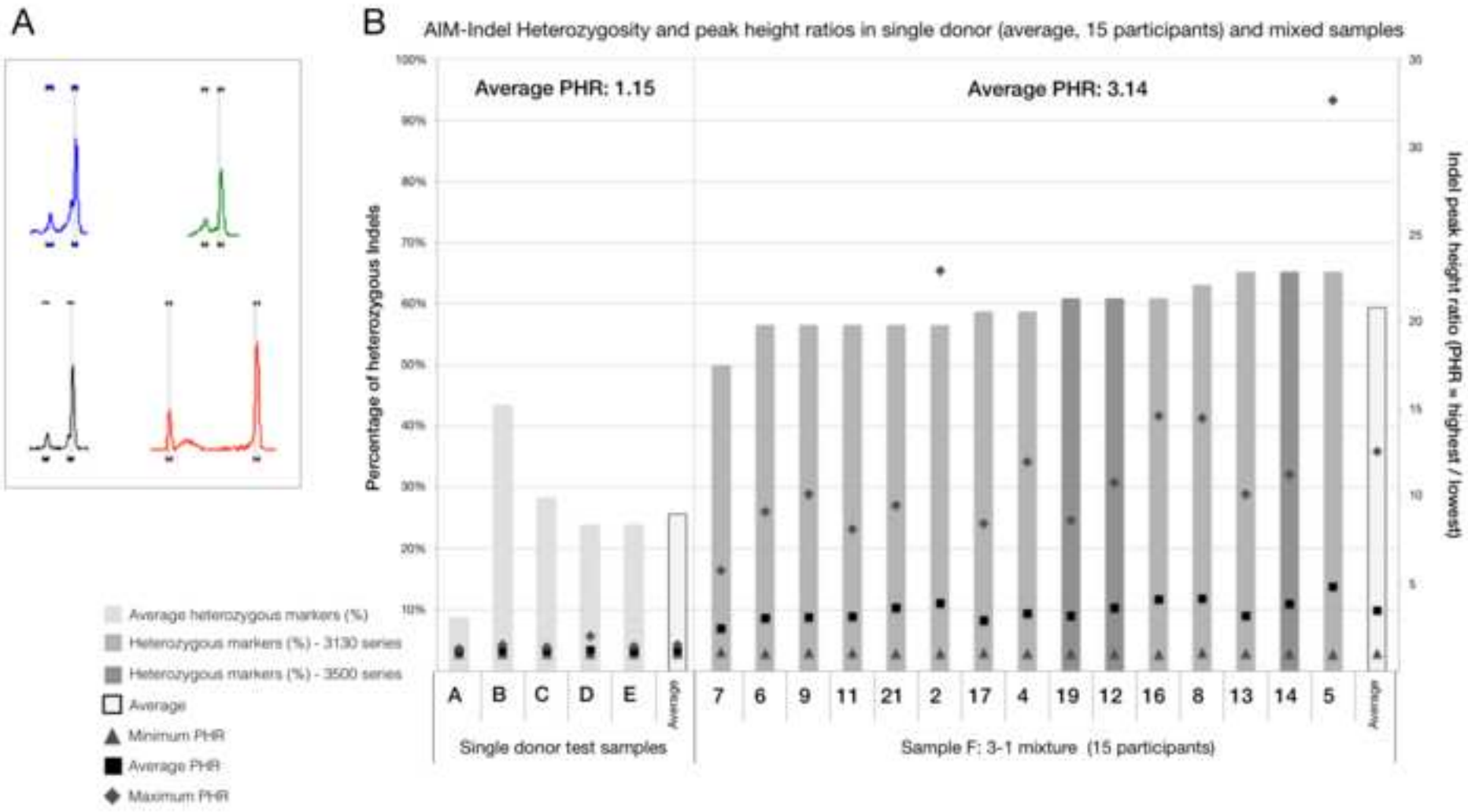


Figure 5

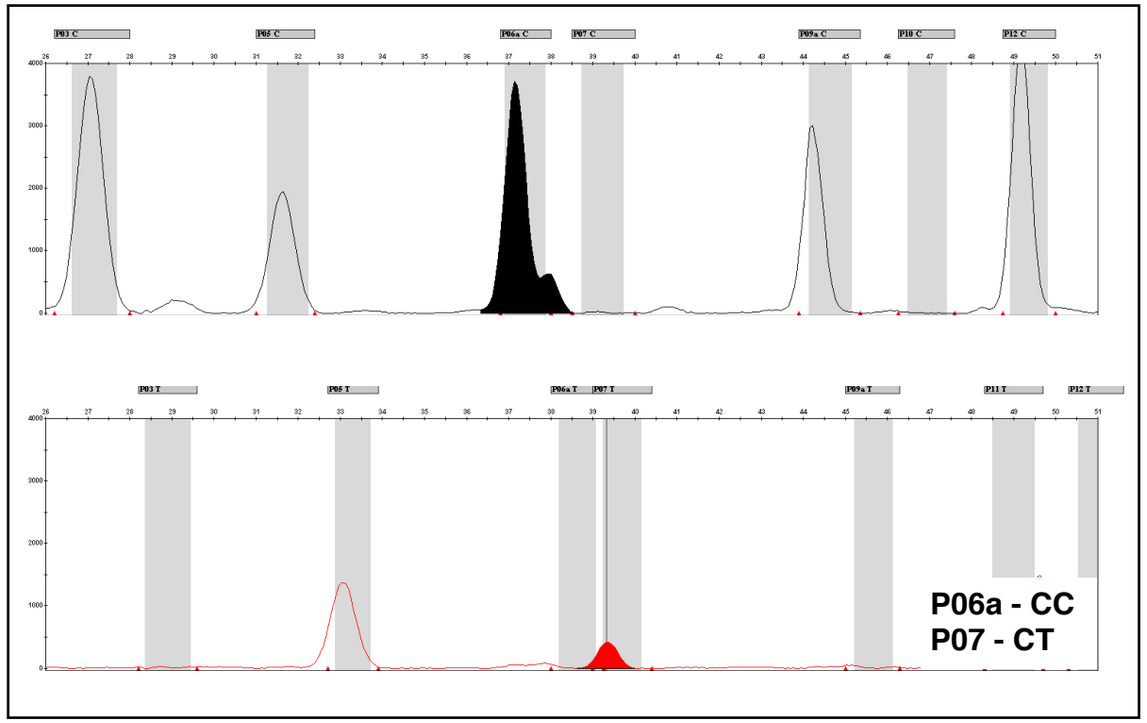
Click here to download high resolution image

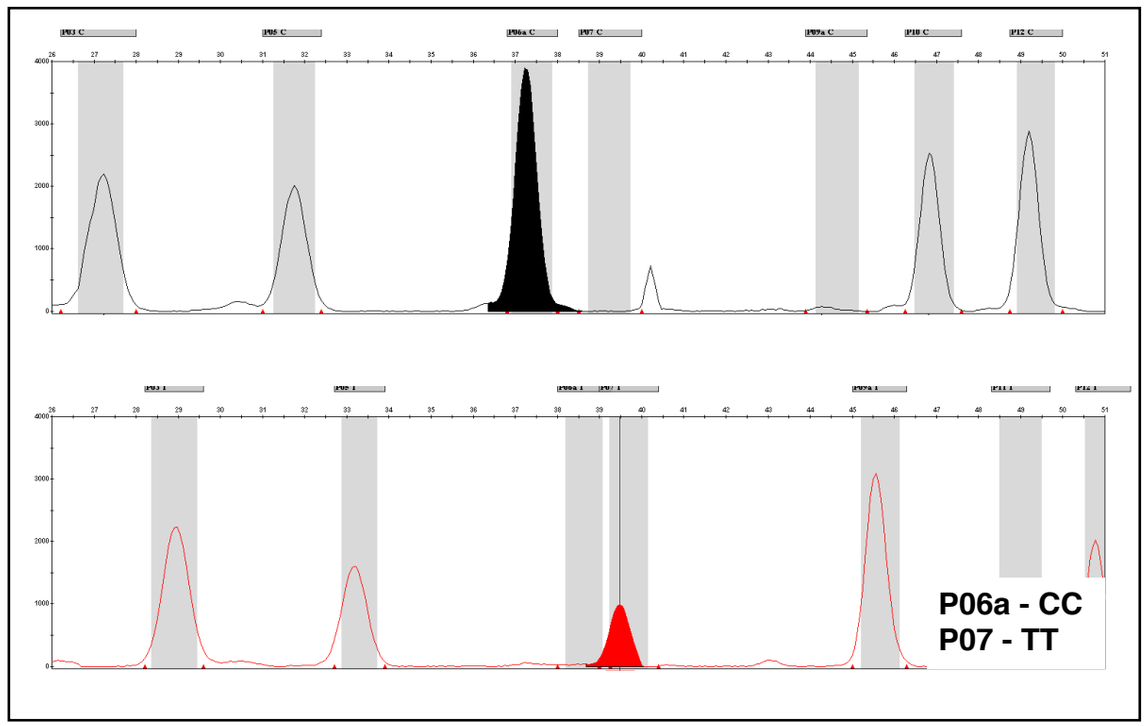


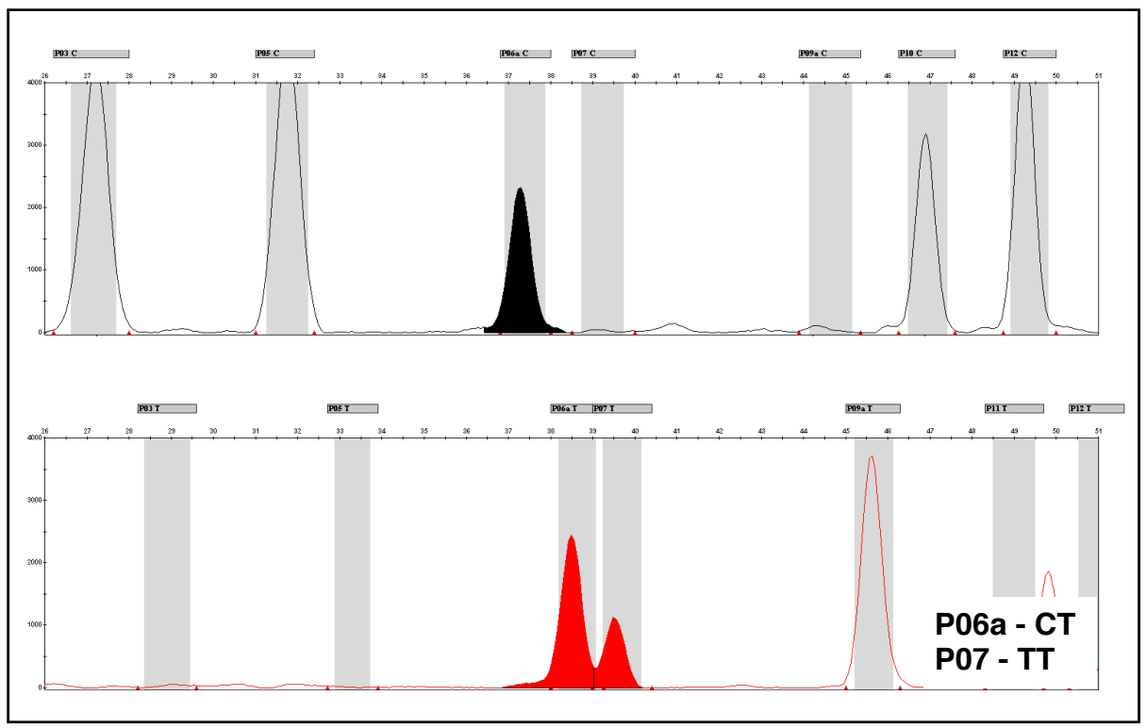
Click here to download e-component: Supplementary Figs S1.pdf

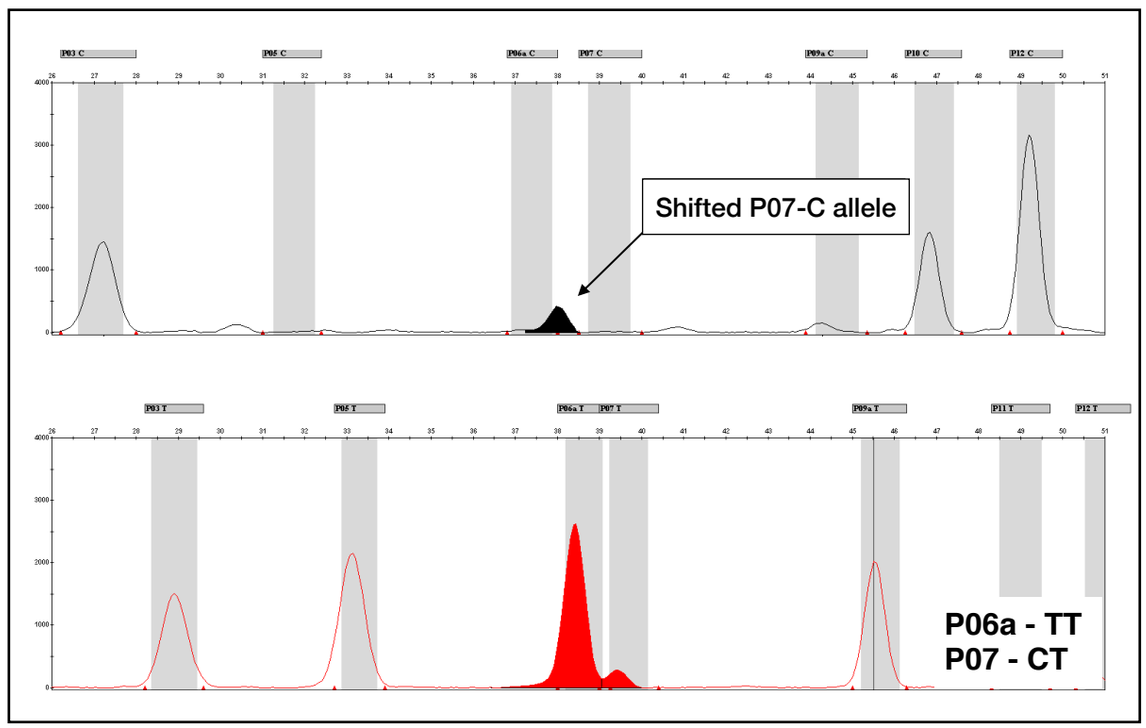
Supplementary Fig. S1 Examples of genotyping challenges in 34-plex or Indel profiles

S1.1 SNPs P06a-P07 (rs10843344-rs239031) peak pairs run very close together

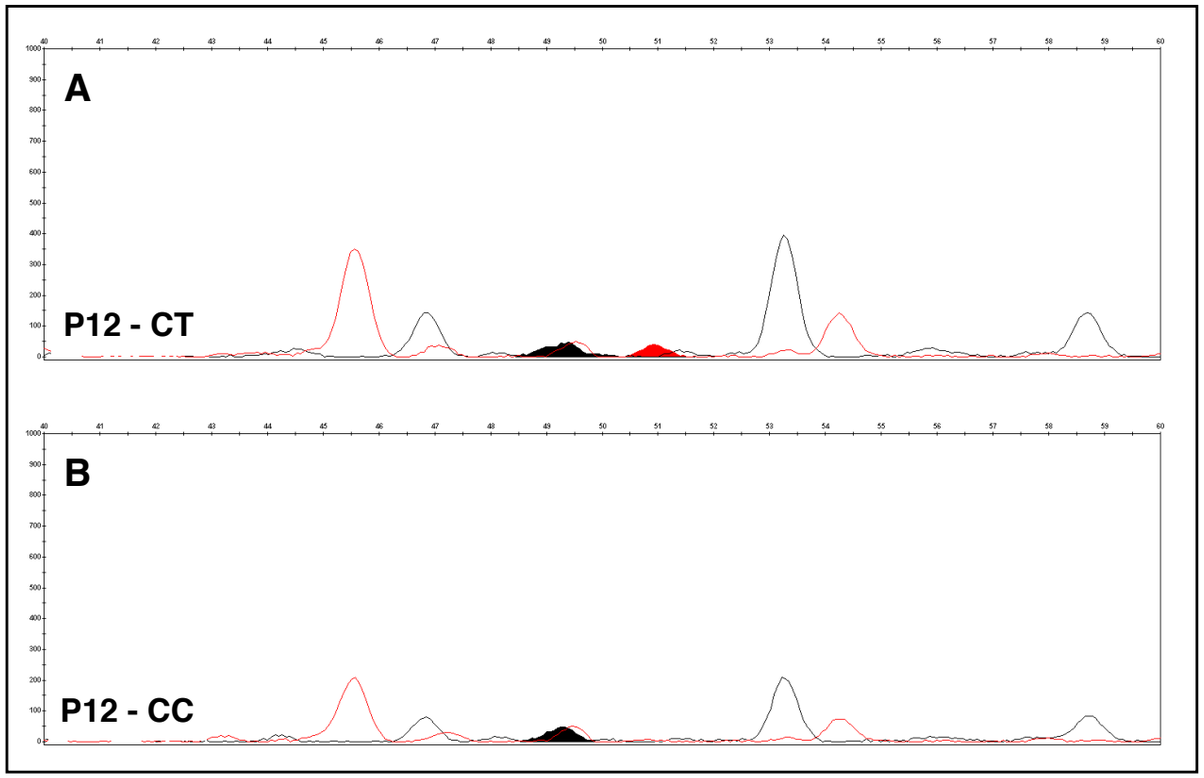


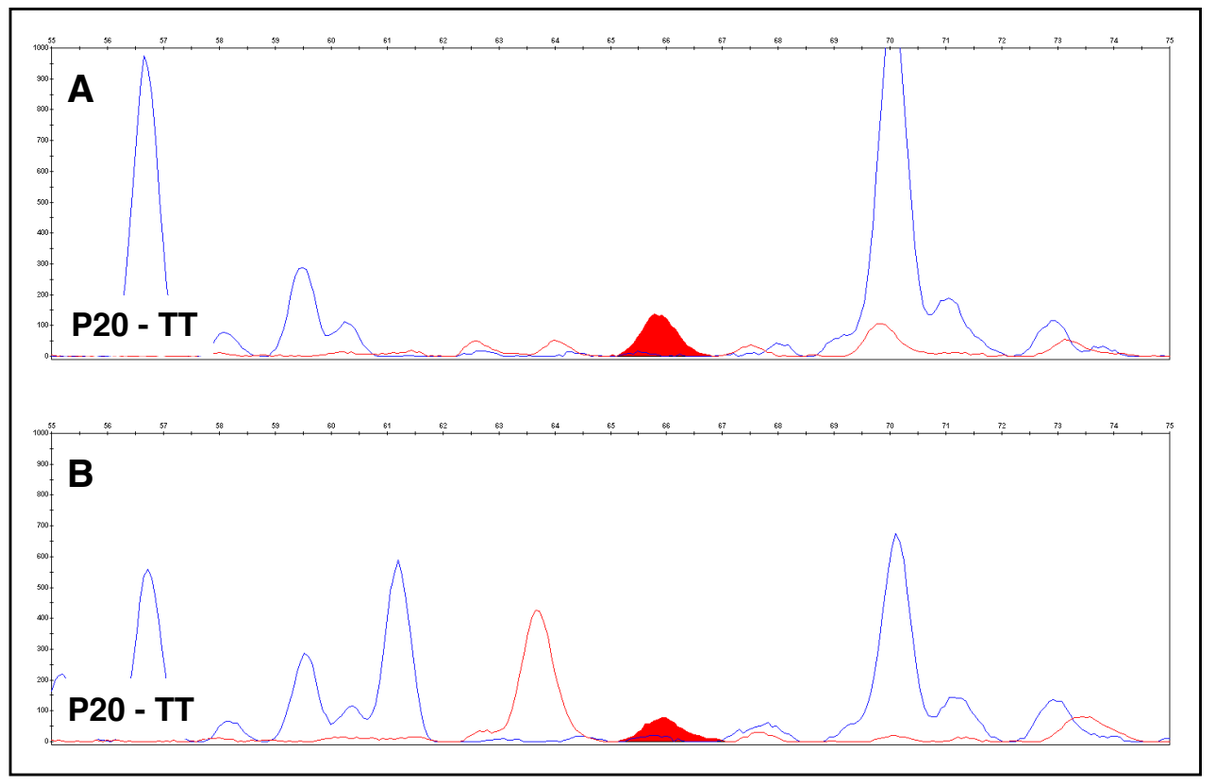


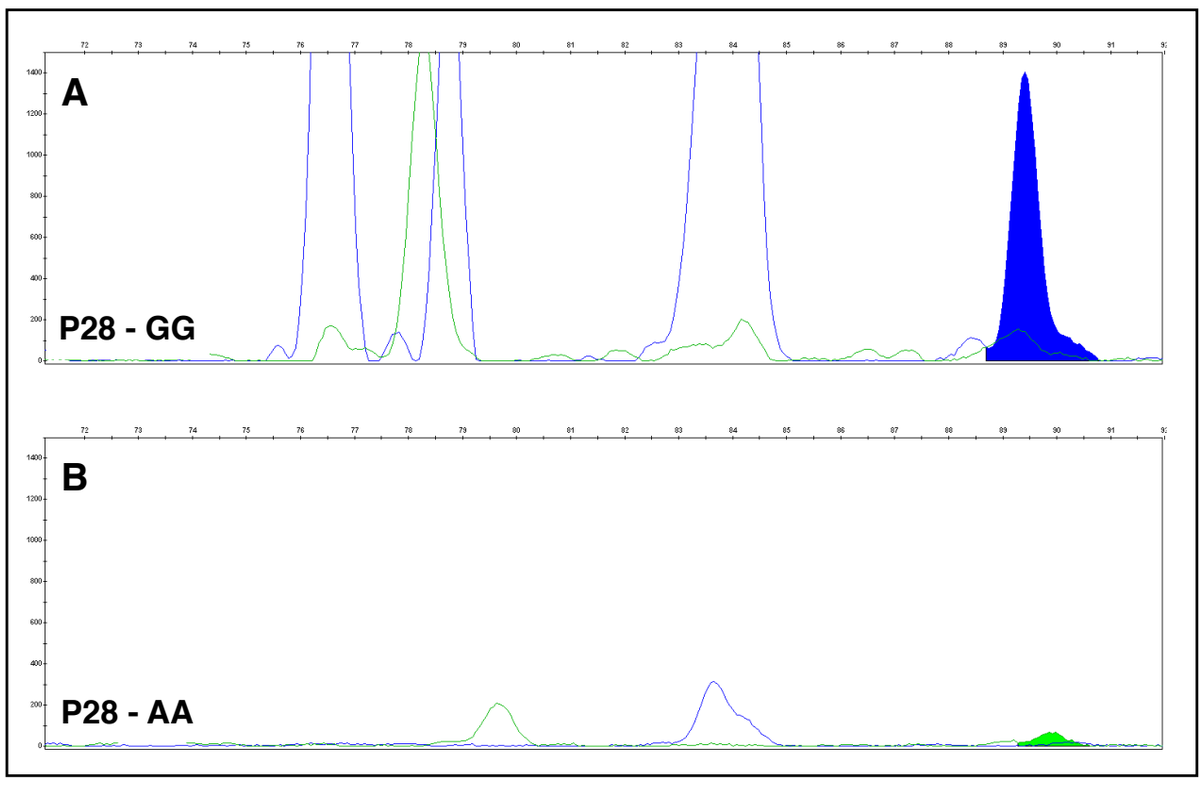




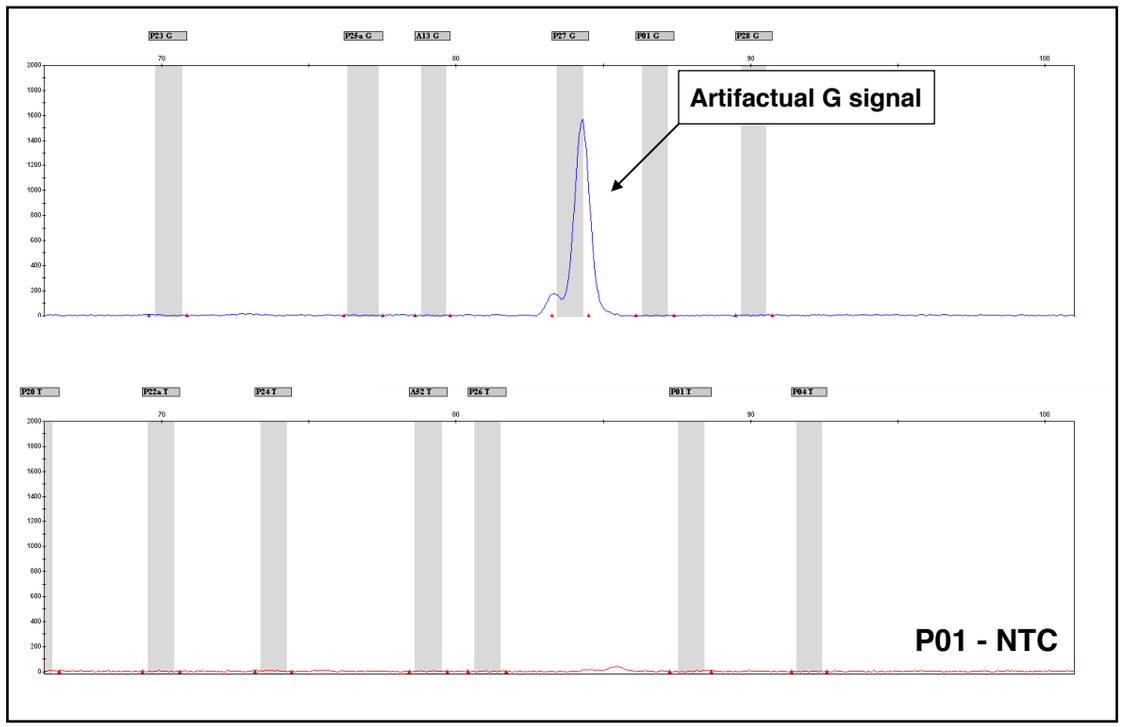
S1.2 SNPs P12/P20/P28 (rs182549/rs1881929/rs3827760) show very low peaks for one or both alleles, particularly allele A in SNP P28.

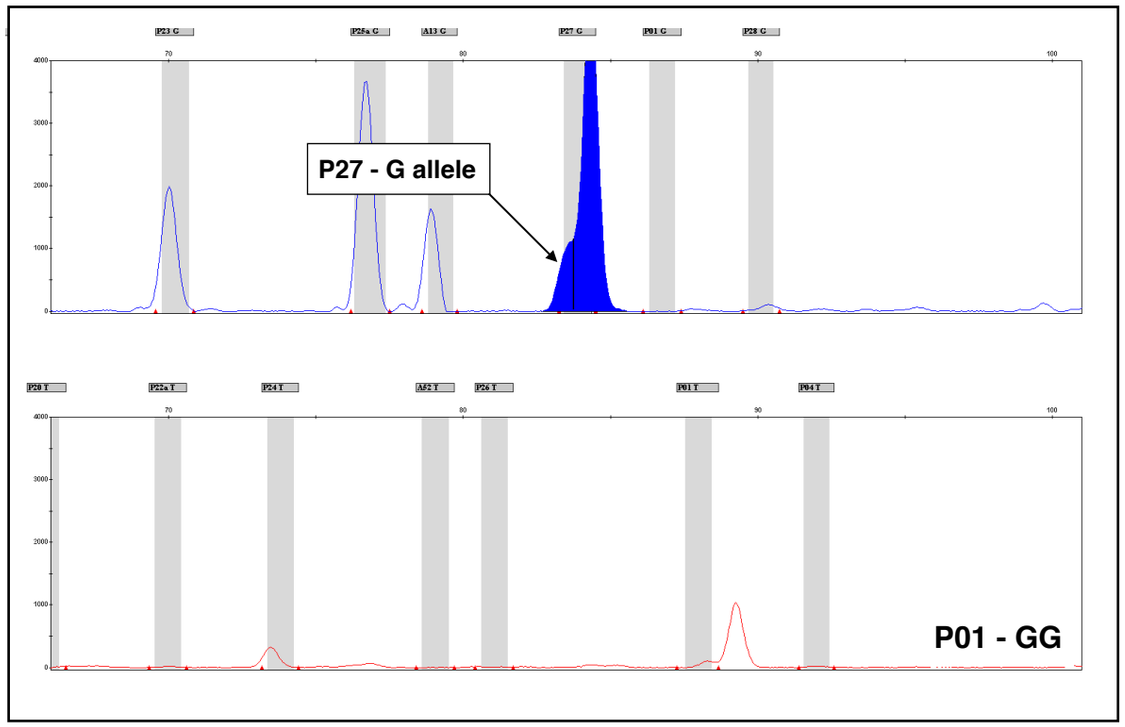


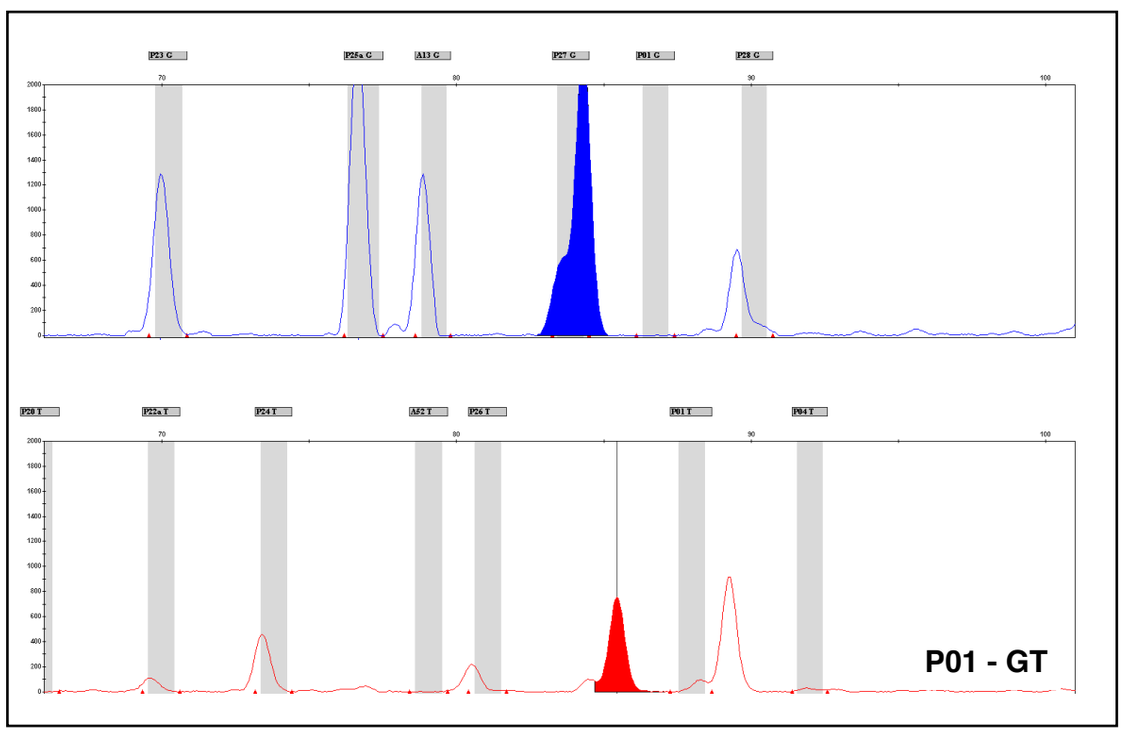


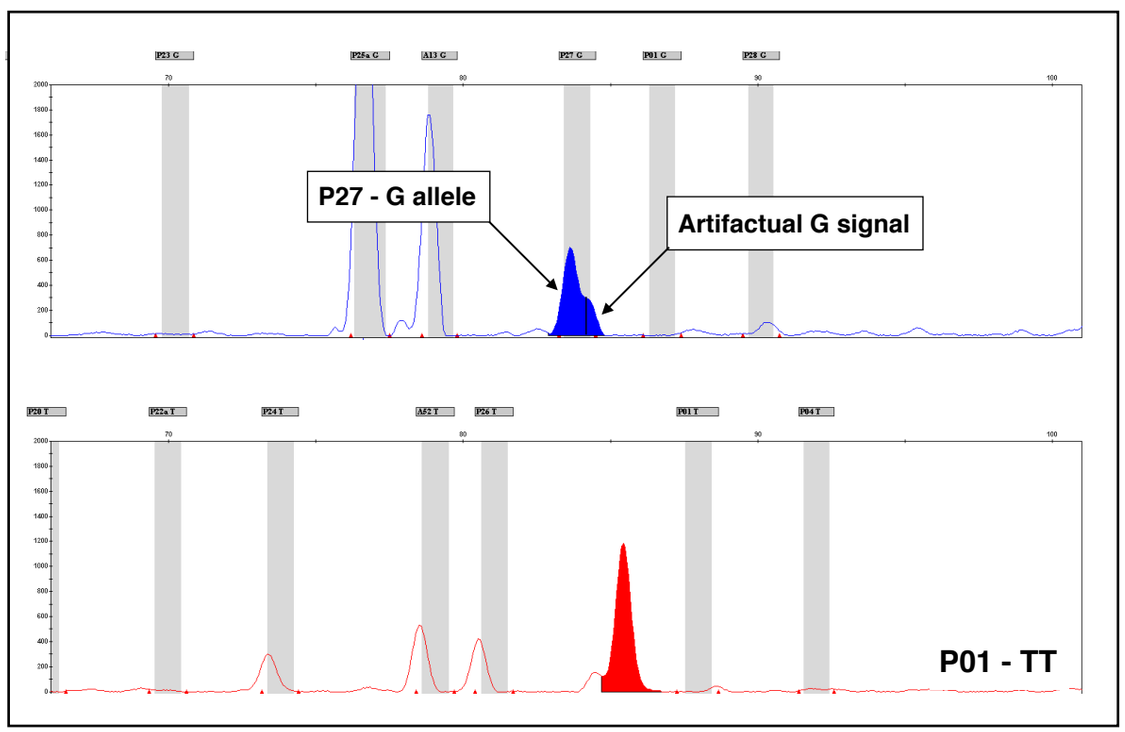


S1.3 P01 (rs2304925) shows an artifactual G signal in the negative control very close to the G peak of P27 (rs5030240-rs2304925). This peak is much higher than the T peak when a true allelic extension product but much lower when artifactual.

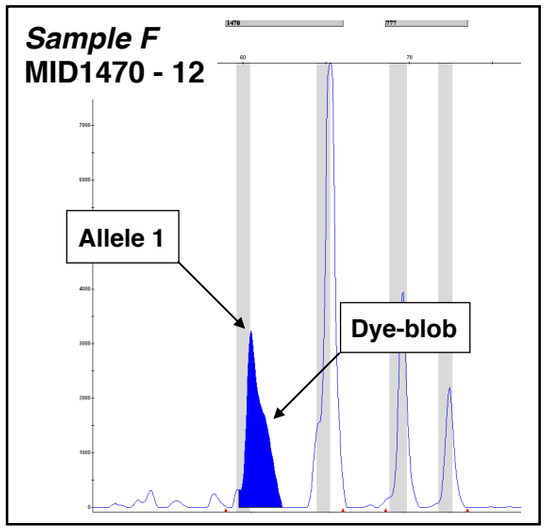


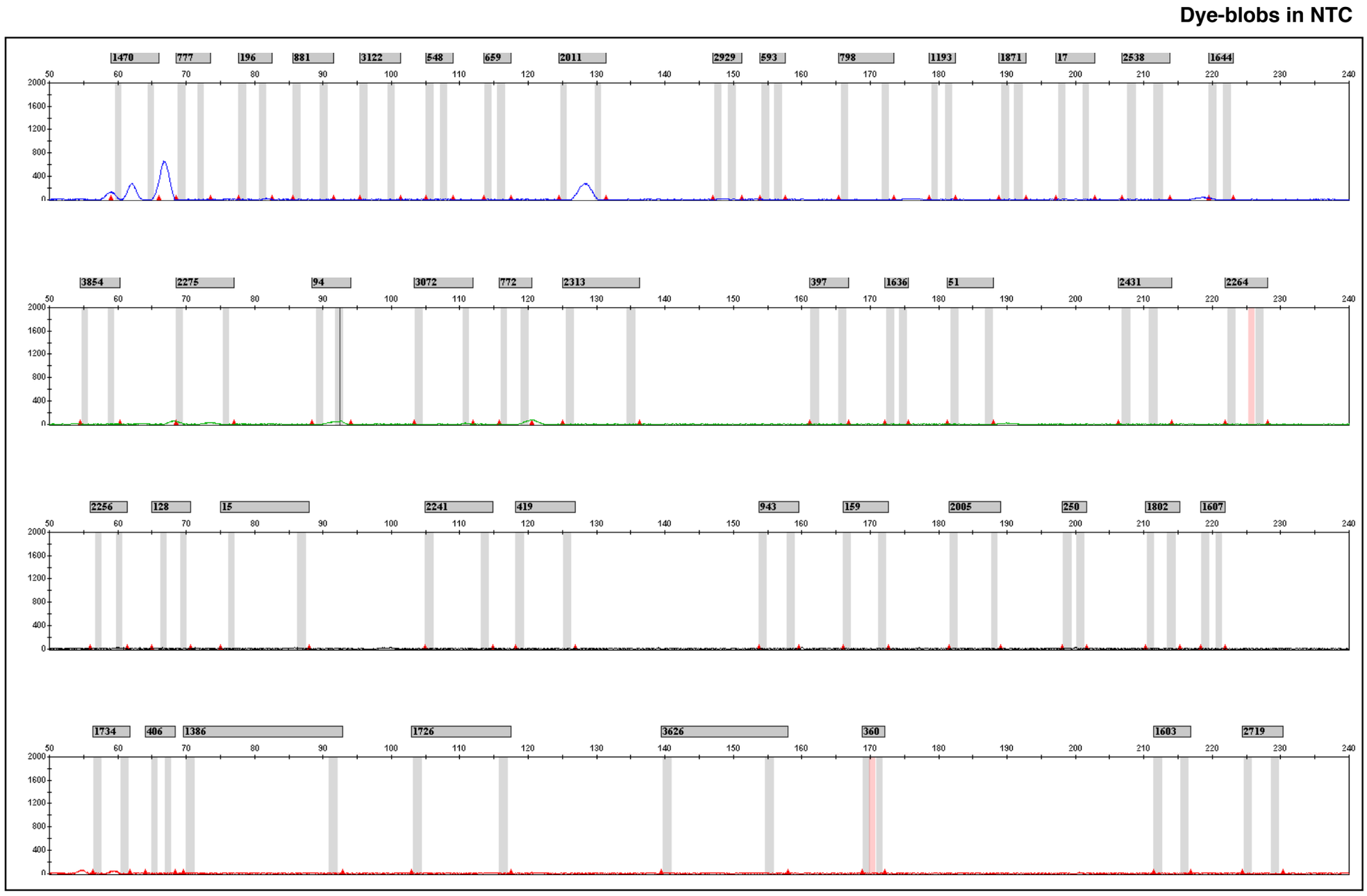




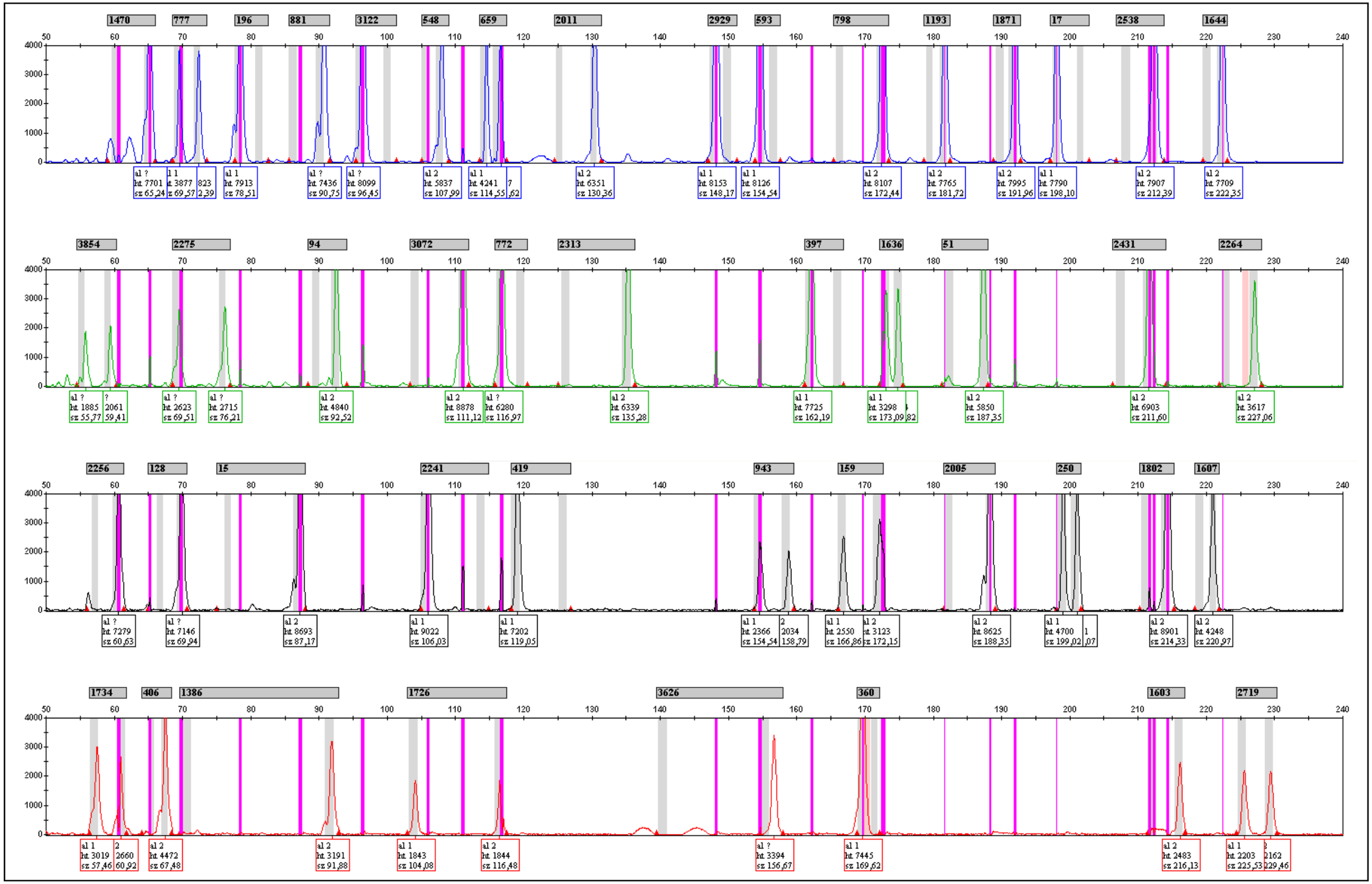


S1.4 Indel dye-blobs present in non-allelic positions in a typical negative template control (NTC) profile. One example of a dye-blob is shown in the inset top-left. This is very close to an allele of the short-amplicon Indel MID1470 (rs2307666), influencing its estimated mid-peak position slightly.

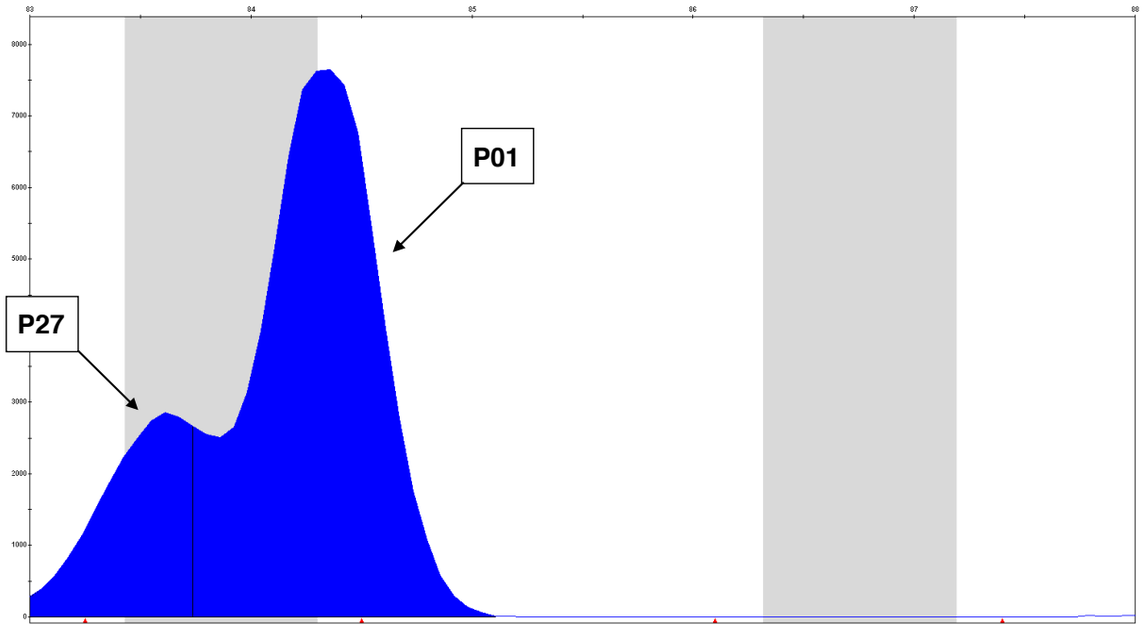


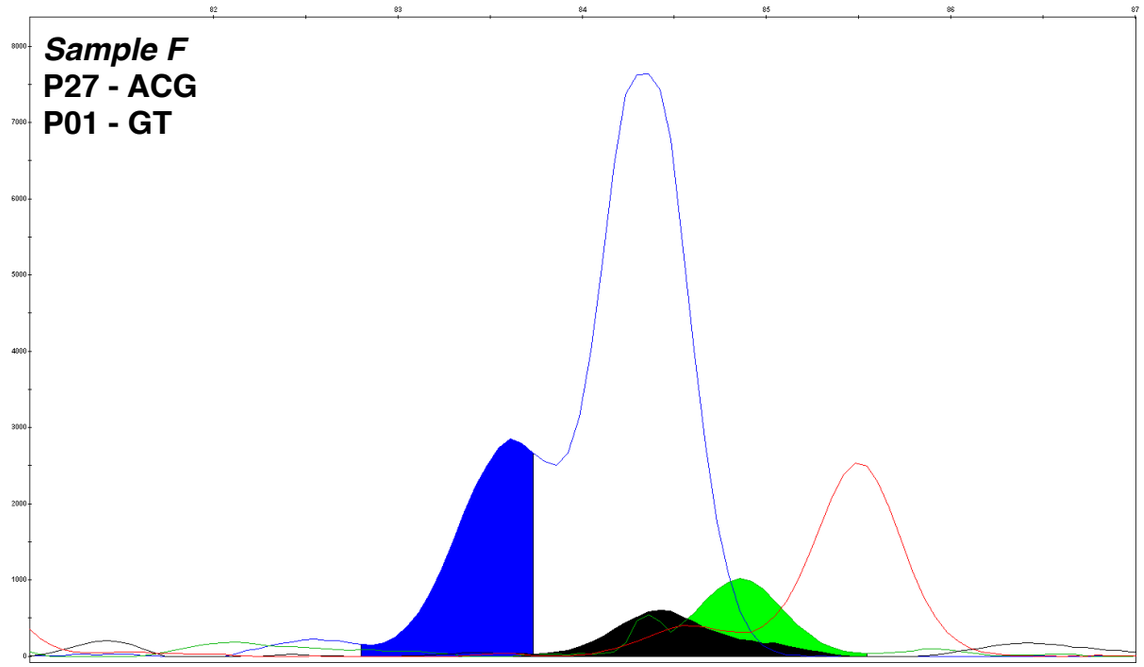


S1.5 Example of strong signal pull-up in an Indel profile due to an overloaded sample



S1.6 SNP P27 (rs5030240) is a tri-allelic marker that showed three alleles in mixed sample F but has low relative peak height for the C allele compared to those of A and G alleles. Peak patterns shown left give another example of the close peak positions shown in S1.3.





Supplementary Table S1
Click here to download e-component: Supplementary Table S1.docx

Supplementary Table S1. Capillary electrophoresis (CE) details for participating laboratories. Grey bars for 34-plex denote five participants not completing SNP genotyping with SNaPshot.

Lab.	CE Detector	Polymer	Dilution factor	
			34-plex*	AIM-Indels
1	3130xl	POP-4	None	None
2	3130xl	POP-4	None	1:10 (A-E) 1:5 (F & NTC)
4	3100	POP-6	None	None
5	3130xl	POP-4		None
6	3130xl	POP-7	None	1:10
7	3130	POP-4	None	None
8	3130xl	POP-4	None	1:20
9	3130xl	POP-7	None	None
11	3130xl	POP-4	None	1:10 (E & F)
12	3500	POP-4		None
13	3130	POP-4	1:10	1:10
14	3500xl	POP-4		1:5
15	3130xl	POP-4		(Not reported)
16	3130xl	POP-4		None
17	3130xl	POP-4	None	1:20
18	3130	POP-4	None	None
19	3500xl	POP-4	None	1:20
20	3130xl	POP-4	None	1:10
21	3130	POP-4	None	None

* 5/19 laboratories marked did not run the 34-plex SNP assay, but this does not show a relationship to the choice of polymer or CE detector used.

Supplementary File S1

Click here to download e-component: Supplementary File S1.xlsx

34-plex AIM-SNPs

1 < µl decimals (1 or 2)

PCR mix:

5 < add prefered % pipetting top-up here

16 < add sample multiple here

	x1 sample	
Buffer 10x	0.69 µl	11.6 µl
BSA (1.6 µg/µl)	0.69 µl	11.6 µl
MgCl2 (25 mM)	1.63 µl	27.4 µl
dNTPs (10 mM)	0.43 µl	7.2 µl
PCR primer mix	1 µl	16.8 µl
AmpliTaq Gold	0.1 µl	1.7 µl
H_2O	0.9 µl	14.4 µl

11.59 11.60
1.5 1.50
27.38 27.40
7.22 7.20
1.50
16.80 16.80
1.50
1.68 1.70
0.86 0.90 14.45
5.40 5.40

0.5 < add DNA concentratio

5.4 µl Mix Total Volume + **Optimum DNA input is 0.75 ng** 1.5 µl = **6.9 µl Total Volume**

Exo-SAP purification:

low-cost

	x1 sample	x1 sample (non evidential DNA)
ExoSAPit	1.3 µl	0.65 µl
PCR product	2.5 µl	1.25 µl

EXT mix:

x1 sample 16 x

Supplementary Files S2.1

Click here to download e-component: AIM-indelplex_POP4_bins.txt

Version	GM v 3.0		
Chemistry Kit	AIM-indelplex		
BinSet Name	AIM-indelplex		
Panel Name	AIM-indelplex		
Marker Name	1470		
1	60.17	0.5	0.76
2	65.1	0.5	0.5
Marker Name	777		
1	69.33	0.5	0.5
2	72.33	0.5	0.5
Marker Name	196		
1	78.37	0.5	0.5
2	81.37	0.5	0.5
Marker Name	881		
1	86.37	0.5	0.5
2	90.37	0.5	0.5
Marker Name	3122		
1	96.18	0.5	0.5
2	100.18	0.5	0.5
Marker Name	548		
1	105.8	0.5	0.5
2	107.8	0.5	0.5
Marker Name	659		
1	114.3	0.5	0.5
2	116.3	0.5	0.5
Marker Name	2011		
1	125.29	0.5	0.5
2	130.29	0.5	0.5
Marker Name	2929		
1	147.91	0.5	0.5
2	149.91	0.5	0.5
Marker Name	593		
1	154.56	0.5	0.5
2	156.56	0.5	0.5
Marker Name	798		
1	166.38	0.5	0.5
2	172.28	0.5	0.5
Marker Name	1193		
1	179.47	0.5	0.5
2	181.47	0.5	0.5
Marker Name	1871		
1	189.75	0.5	0.5
2	191.75	0.5	0.66
Marker Name	17		
1	197.81	0.5	0.5
2	201.81	0.5	0.5
Marker Name	2538		
1	208.38	0.5	0.75
2	212.38	0.5	0.75
Marker Name	1644		
1	220.2	0.5	0.5
2	222.2	0.5	0.5

Marker Name	3854			
1	55.4	0.5	0.5	
2	59.21	0.59	0.5	
Marker Name	2275			
1	69.3	0.5	0.5	
2	76.09	0.5	0.5	
Marker Name	3072			
1	104.07	0.5	0.5	
2	111.07	0.5	0.5	
Marker Name	772			
1	116.6	0.5	0.5	
2	119.6	0.5	0.5	
Marker Name	2313			
1	126.01	0.5	0.5	
2	135.01	0.5	0.5	
Marker Name	397			
1	161.99	0.5	0.5	
2	165.99	0.5	0.5	
Marker Name	1636			
1	172.86	0.5	0.5	
2	174.86	0.5	0.5	
Marker Name	51			
1	182.11	0.5	0.5	
2	187.11	0.5	0.5	
Marker Name	2431			
1	207.47	0.5	0.75	
2	211.47	0.5	0.75	
Marker Name	2264			
1	222.8	0.5	0.5	
3	225.8	0.4	0.4	mutant
2	226.8	0.4	0.66	
Marker Name	2256			
1	57.38	0.5	0.5	
2	60.38	0.5	0.5	
Marker Name	128			
1	66.67	0.5	0.5	
2	69.67	0.5	0.5	
Marker Name	15			
1	76.1	0.5	0.5	
2	87.0	0.5	0.64	
Marker Name	2241			
1	106.0	0.64	0.5	
2	114.01	0.5	0.5	
Marker Name	419			
1	118.84	0.5	0.65	
2	125.84	0.5	0.5	
Marker Name	943			
1	154.51	0.56	0.5	
2	158.51	0.5	0.5	
Marker Name	159			
1	166.85	0.61	0.5	
2	171.9	0.5	0.5	

Marker Name	2005			
1	182.28	0.5	0.55	
2	188.28	0.64	0.5	
Marker Name	250			
1	198.85	0.5	0.5	
2	200.85	0.5	0.5	
Marker Name	1802			
1	211.13	0.5	0.5	
2	214.13	0.5	0.5	
Marker Name	1607			
1	219.02	0.5	0.5	
2	220.95	0.62	0.5	
Marker Name	406			
1	65.39	0.5	0.5	
2	67.39	0.5	0.5	
Marker Name	1386			
1	70.71	0.5	0.71	
2	91.57	0.5	0.5	
Marker Name	1726			
1	103.82	0.5	0.5	
2	116.5	0.5	0.5	
Marker Name	3626			
1	140.36	0.5	0.5	
2	156.25	0.5	0.5	
Marker Name	360			
1	169.42	0.4	0.4	
3	170.42	0.4	0.4	mutant
2	171.42	0.4	0.4	
Marker Name	1603			
1	212.05	0.5	0.56	
2	215.97	0.5	0.5	
Marker Name	2719			
1	225.38	0.5	0.5	
2	229.38	0.5	0.5	
Marker Name	1734			
1	57.11	0.5	0.5	
2	60.79	0.5	0.5	
Marker Name	94			
1	89.14	0.5	0.5	
2	92.14	0.5	0.56	

Supplementary Files S2.2

Click here to download e-component: AIM-indelplex_POP4_Panels.txt

Version GM v 3.0
Kit type: MICROSATELLITE
Chemistry Kit AIM-indelplex none
Panel AIM-indelplex none

1470	blue	59.498930343000005	65.81825256900001	-	9	0.0	none	
777	blue	68.55843072	73.031835852	-	9	0.0	none	
196	blue	77.588872041	82.088429436	-	9	0.0	none	
881	blue	85.617309136	91.15076970400001	-	9	0.0	none	
3122	blue	95.320690594	100.900499185	-	9	0.0	none	
548	blue	105.057912973	108.532867596	-	9	0.0	none	
659	blue	113.556264981	117.080076701	-	9	0.0	none	
2011	blue	124.498164706	131.041303657	-	9	0.0	none	
2929	blue	146.955519673	150.824584052	-	9	0.0	none	
593	blue	153.83022589700002	157.14313298399998	-	9	0.0	none	
798	blue	165.32897286899998	173.501119818	-	9	0.0	none	
1193	blue	178.658800433	182.45524446299999	-	9	0.0	none	
1871	blue	188.823452509	192.72466145	-	9	0.0	none	
17	blue	197.084416798	202.680731973	-	9	0.0	none	
2538	blue	206.773228342	213.827474758	-	9	0.0	none	
1644	blue	219.483987195	223.06873612500002	-	9	0.0	none	
3854	Green	54.61523001	59.995735289	-	9	0.0	none	
2275	Green	68.507204408	76.863106493	-	9	0.0	none	
3072	Green	103.268780394	111.960791133	-	9	0.0	none	
772	Green	115.766532729	120.46989852499999	-	9	0.0	none	
2313	Green	125.024081548	136.274250832	-	9	0.0	none	
397	Green	161.189615282	166.818202663	-	9	0.0	none	
1636	Green	172.08629923200002	175.600073663	-	9	0.0	none	
51	Green	181.211398852	188.025739155	-	9	0.0	none	
2431	Green	206.293870962	214.07491917299998	-	9	0.0	none	
2264	Green	221.926816896	228.05683674	-	9	0.0	none	
2256	Yellow	56.680080642	61.154696174	-	9	0.0	none	
128	Yellow	65.93894781700001	70.400997293	-	9	0.0	none	
15	Yellow	75.5	87.975809403	-	9	0.0	none	
2241	Yellow	104.973806014	114.835216709	-	9	0.0	none	
419	Yellow	118.07762833899999	126.850206737	-	9	0.0	none	
943	Yellow	153.696434559	159.341659618	-	9	0.0	none	
159	Yellow	166.02409311300002	172.70691387	-	9	0.0	none	
2005	Yellow	181.501371715	189.06208221400001	-	9	0.0	none	
250	Yellow	198.065914853	201.572590476	-	9	0.0	none	
1802	Yellow	210.217316797	215.228136625	-	9	0.0	none	
1607	Yellow	218.273123059	221.87207210900002	-	9	0.0	none	
406	Red	64.678876207	68.176704628	-	9	0.0	none	
1386	Red	69.56966599	92.812677351	-	9	0.0	none	
1726	Red	102.977336125	117.29484620400001	-	9	0.0	none	
3626	Red	139.43282453700002	156.750843119	-	9	0.0	none	
360	Red	168.792202855	172.14353350399998	-	9	0.0	none	
1603	Red	211.384383599	216.862882022	-	9	0.0	none	
2719	Red	224.388680709	230.34325627	-	9	0.0	none	
1734	Red	56.285110841	61.814691114000006	-	9	0.0	none	
94	Green	88.363699841	92.99304501099999	-	9	0.0	none	

Supplementary Files S2.3

Click here to download e-component: 34plex_POP4__bins.txt

```
Version GM v 3.0
Chemistry Kit 34plex_POP4
BinSet Name 34plex_POP4
Panel Name 34plex_POP4
Marker Name P01 T
T 85.75 0.4 0.4
Marker Name A07 G
G 26.92 0.53 0.48
Marker Name A07 A
A 29.26 0.46 0.42
Marker Name P03 C
C 27.22 0.6 0.47
Marker Name P03 T
T 28.98 0.62 0.46
Marker Name P04 C
C 88.76 0.4 0.4
Marker Name P04 T
T 89.56 0.4 0.4
Marker Name A29 G
G 28.44 0.46 0.45
Marker Name A29 A
A 30.49 0.49 0.47
Marker Name P05 C
C 31.79 0.53 0.45
Marker Name P05 T
T 33.3 0.42 0.41
Marker Name A21 G
G 34.18 0.4 0.41
Marker Name A21 A
A 36.46 0.51 0.47
Marker Name P06a C
C 37.39 0.48 0.47
Marker Name P06a T
T 38.65 0.45 0.42
Marker Name P08 G
G 40.37 0.45 0.41
Marker Name P08 A
A 41.01 0.41 0.42
Marker Name P07 C
C 39.23 0.5 0.5
Marker Name P07 T
T 39.67 0.43 0.47
Marker Name A40 G
G 43.22 0.42 0.42
Marker Name A40 A
A 44.34 0.51 0.41
Marker Name P09a C
C 44.64 0.5 0.5
Marker Name P09a T
T 45.67 0.46 0.44
Marker Name P10 G
G 46.52 0.42 0.42
```

Marker Name P10 C
C 46.94 0.46 0.46
Marker Name P11 A
A 49.0 0.5 0.47
Marker Name P11 T
T 49.63 0.4 0.4
Marker Name P12 C
C 49.34 0.43 0.46
Marker Name P12 T
T 50.97 0.44 0.41
Marker Name P13 G
G 49.85 0.5 0.5
Marker Name P02 A
A 88.17 0.5 0.5
Marker Name P02 C
C 87.87 0.46 0.45
Marker Name P01 G
G 84.55 0.4 0.4
Marker Name P13 A
A 51.5 0.46 0.45
Marker Name P14 C
C 53.37 0.43 0.46
Marker Name P14 T
T 54.34 0.4 0.41
Marker Name P15 G
G 55.29 0.47 0.45
Marker Name P15 A
A 56.1 0.47 0.43
Marker Name P16a G
G 56.79 0.42 0.45
Marker Name P16a A
A 57.64 0.4 0.4
Marker Name P17 C
C 58.84 0.44 0.42
Marker Name P17 T
T 59.56 0.44 0.43
Marker Name P18 G
G 61.26 0.41 0.42
Marker Name P18 A
A 62.14 0.47 0.41
Marker Name P19 C
C 62.57 0.46 0.41
Marker Name P19 T
T 63.82 0.4 0.4
Marker Name P20 G
G 63.87 0.45 0.42
Marker Name P20 T
T 65.82 0.43 0.45
Marker Name P21 A
A 66.17 0.47 0.43
Marker Name P21 C
C 66.01 0.46 0.43

Marker Name	P22a C		
C	69.27	0.4	0.4
Marker Name	P22a T		
T	69.98	0.44	0.43
Marker Name	P23 G		
G	70.23	0.45	0.45
Marker Name	P23 A		
A	70.24	0.5	0.5
Marker Name	P24 A		
A	73.53	0.47	0.4
Marker Name	P24 C		
C	73.08	0.4	0.4
Marker Name	P24 T		
T	73.83	0.46	0.4
Marker Name	A52 A		
A	78.77	0.43	0.48
Marker Name	A52 T		
T	79.01	0.4	0.49
Marker Name	P25a G		
G	76.83	0.51	0.52
Marker Name	P25a C		
C	77.18	0.48	0.46
Marker Name	P26 C		
C	80.5	0.4	0.47
Marker Name	P26 T		
T	81.07	0.44	0.43
Marker Name	A13 G		
G	79.24	0.43	0.4
Marker Name	A13 A		
A	80.07	0.4	0.47
Marker Name	P27 G		
G	83.88	0.45	0.42
Marker Name	P27 A		
A	85.12	0.42	0.42
Marker Name	P27 C		
C	84.71	0.42	0.43
Marker Name	P28 G		
G	89.59	0.4	0.4
Marker Name	P28 A		
A	90.5	0.43	0.45

Supplementary Files S2.4

Click here to download e-component: 34plex_POP4_Panels.txt

Version GM v 3.0
Kit type: MICROSATELLITE
Chemistry Kit 34plex_POP4 none
Panel 34plex_POP4 none

P01 T	Red	85.0	86.5	-	2	0.0	rs2304925	-
A07 G	Blue	25.95	27.5	-	2	0.0	rs917118	-
A07 A	Green	28.5	29.88	-	2	0.0	rs917118	-
P03 C	Yellow	26.2	28.0	-	2	0.0	rs1321333	-
P03 T	Red	28.2	29.6	-	2	0.0	rs1321333	-
P04 C	Yellow	88.1	89.5	-	2	0.0	rs2814778	-
P04 T	Red	88.7	90.5	-	2	0.0	rs2814778	-
A29 G	Blue	27.8	29.17	-	2	0.0	rs1024116	-
A29 A	Green	29.8	31.1	-	2	0.0	rs1024116	-
P05 C	Yellow	31.0	32.4	-	2	0.0	rs7897550	-
P05 T	Red	32.7	33.9	-	2	0.0	rs7897550	-
A21 G	Blue	33.6	34.71	-	2	0.0	rs722098	-
A21 A	Green	35.75	37.1	-	2	0.0	rs722098	-
P06a C	Yellow	36.8	38.0	-	2	0.0	rs10843344	-
P06a T	Red	38.0	39.25	-	2	0.0	rs10843344	-
P08 G	Blue	39.8	41.01	-	2	0.0	rs12913832	-
P08 A	Green	40.4	41.65	-	2	0.0	rs12913832	-
P07 C	Yellow	38.5	40.0	-	2	0.0	rs239031	-
P07 T	Red	39.0	40.4	-	2	0.0	rs239031	-
A40 G	Blue	42.6	43.85	-	2	0.0	rs2040411	-
A40 A	Green	43.6	44.93	-	2	0.0	rs2040411	-
P09a C	Yellow	43.9	45.35	-	2	0.0	rs1978806	-
P09a T	Red	45.0	46.3	-	2	0.0	rs1978806	-
P10 G	Blue	45.9	47.2	-	2	0.0	rs773658	-
P10 C	Yellow	46.25	47.6	-	2	0.0	rs773658	-
P11 A	Green	48.3	49.7	-	2	0.0	rs10141763	-
P11 T	Red	48.8	50.5	-	2	0.0	rs10141763	-
P12 C	Yellow	48.75	50.0	-	2	0.0	rs182549	-
P12 T	Red	50.3	51.6	-	2	0.0	rs182549	-
P13 G	Blue	49.1	50.7	-	2	0.0	rs1573020	-
P02 A	Green	87.5	88.85	-	2	0.0	rs5997008	-
P02 C	Yellow	87.19	88.5	-	2	0.0	rs5997008	-
P01 G	Blue	84.1	85.3	-	2	0.0	rs2304925	-
P13 A	Green	50.85	52.2	-	2	0.0	rs1573020	-
P14 C	Yellow	52.75	54.0	-	2	0.0	rs896788	-
P14 T	Red	53.75	54.95	-	2	0.0	rs896788	-
P15 G	Blue	54.65	55.9	-	2	0.0	rs2065160	-
P15 A	Green	55.4	56.75	-	2	0.0	rs2065160	-
P16a G	Blue	56.2	57.45	-	2	0.0	rs2572307	-
P16a A	Green	56.7	58.2	-	2	0.0	rs2572307	-
P17 C	Yellow	58.2	59.4	-	2	0.0	rs2303798	-
P17 T	Red	59.0	60.2	-	2	0.0	rs2303798	-
P18 G	Blue	60.65	61.9	-	2	0.0	rs2065982	-
P18 A	Green	61.5	62.75	-	2	0.0	rs2065982	-
P19 C	Yellow	62.0	63.2	-	2	0.0	rs3785181	-
P19 T	Red	63.22	64.4	-	2	0.0	rs3785181	-
P20 G	Blue	63.22	64.55	-	2	0.0	rs881929	-
P20 T	Red	65.2	66.5	-	2	0.0	rs881929	-

P21 A	Green	65.5	66.8	-	2	0.0	rs1498444	-
P21 C	Yellow	65.45	66.6	-	2	0.0	rs1498444	-
P22a C	Yellow	68.65	69.85	-	2	0.0	rs1426654	-
P22a T	Red	69.35	70.6	-	2	0.0	rs1426654	-
P23 G	Blue	69.55	70.85	-	2	0.0	rs2026721	-
P23 A	Green	69.55	70.95	-	2	0.0	rs2026721	-
P24 A	Green	72.85	74.15	-	2	0.0	rs4540055	-
P24 C	Yellow	72.5	73.65	-	2	0.0	rs4540055	-
P24 T	Red	73.17	74.4	-	2	0.0	rs4540055	-
A52 A	Green	78.15	79.4	-	2	0.0	rs1335873	-
A52 T	Red	78.4	79.7	-	2	0.0	rs1335873	-
P25a G	Blue	76.2	77.5	-	2	0.0	rs16891982	-
P25a C	Yellow	76.5	77.8	-	2	0.0	rs16891982	-
P26 C	Yellow	79.9	81.15	-	2	0.0	rs730570	-
P26 T	Red	80.4	81.7	-	2	0.0	rs730570	-
A13 G	Blue	78.6	79.8	-	2	0.0	rs1886510	-
A13 A	Green	79.5	80.75	-	2	0.0	rs1886510	-
P27 G	Blue	83.25	84.5	-	2	0.0	rs5030240	-
P27 A	Green	84.5	85.75	-	2	0.0	rs5030240	-
P27 C	Yellow	84.1	85.35	-	2	0.0	rs5030240	-
P28 G	Blue	89.0	90.5	-	2	0.0	rs3827760	-
P28 A	Green	89.9	91.15	-	2	0.0	rs3827760	-

Supplementary Files S2.7

Click here to download e-component: 34-PLEX_POP7_bins.txt

```
Version GM v 3.0
Chemistry Kit 34-PLEX
BinSet Name 34-PLEX
Panel Name 34-Plex Electrophoretic Shift
Marker Name 01rs1321333
ASR 34.77 37.61
C 34.77 35.9 Yellow
T 36.39 37.61000000000001 Red
Marker Name 02rs917118
ASR 32.6 35.67
G 32.6 33.809999999999995 Blue
A 34.870000000000005 35.67 Green
Marker Name 03rs1024116
ASR 33.35 36.39
G 33.35 34.54 Blue
A 35.17 36.39 Green
Marker Name 04rs7897550
ASR 37.53 40.05
C 37.53 38.36 Yellow
T 38.96 40.050000000000004 Red
Marker Name 05rs722098
ASR 37.7 40.78
G 37.7 38.5 Blue
A 39.980000000000004 40.78 Green
Marker Name 06rs10843344
ASR 41.68 44.0
C 41.68 42.48 Yellow
T 43.2 44.0 Red
Marker Name 07rs239031
ASR 42.35 45.17
C 42.35 43.15 Yellow
T 44.36 45.17 Red
Marker Name 08rs12913832
ASR 44.16 45.64
G 44.160000000000004 44.96 Blue
A 44.84 45.64 Green
Marker Name 09rs2040411
ASR 46.48 48.49
G 46.480000000000004 47.78 Blue
A 47.690000000000005 48.49 Green
Marker Name 10rs1978806
ASR 48.63 50.43
C 48.629999999999995 49.75 Yellow
T 49.63 50.43 Red
Marker Name 11rs773658
ASR 49.49 51.02
G 49.49 50.29 Blue
C 50.22 51.019999999999996 Yellow
Marker Name 12rs10141763
ASR 52.01 54.21
A 52.010000000000005 53.08 Green
T 52.669999999999995 54.21 Red
```

Marker Name	13rs182549		
ASR	51.8	53.81	
C	51.8	52.92	Yellow
T	53.01	53.809999999999995	Red
Marker Name	14rs1573020		
ASR	53.09	54.94	
G	53.09	54.59	Blue
A	53.69	54.94	Green
Marker Name	15rs896788		
ASR	55.66	57.75	
C	55.66	57.03	Yellow
T	56.900000000000006	57.75	Red
Marker Name	16rs2065160		
ASR	56.85	58.04	
G	56.849999999999994	57.75	Blue
A	57.24	58.04	Green
Marker Name	17rs2572307		
ASR	58.03	59.66	
G	58.03	58.83	Blue
A	58.86	59.66	Green
Marker Name	18rs2303798		
ASR	59.78	61.12	
C	59.78	60.58	Yellow
T	60.32	61.12	Red
Marker Name	19rs2065982		
ASR	62.31	63.86	
G	62.31	63.11	Blue
A	63.06	63.86	Green
Marker Name	20rs3785181		
ASR	63.89	66.12	
C	63.89	64.79	Yellow
T	65.32	66.12	Red
Marker Name	21rs881929		
ASR	64.29	67.39	
G	64.29	65.26	Blue
T	66.58999999999999	67.39	Red
Marker Name	22rs1498444		
ASR	67.19	68.17	
A	67.19	67.99000000000001	Green
C	67.36999999999999	68.17	Yellow
Marker Name	23rs1426654		
ASR	70.48	72.13	
C	70.47999999999999	71.28	Yellow
T	71.33	72.13000000000001	Red
Marker Name	24rs2026721		
ASR	71.28	72.65	
G	71.28	72.08000000000001	Blue
A	71.85	72.65	Green
Marker Name	25rs4540055		
ASR	74.18	76.08	
A	74.5	75.30000000000001	Green
C	74.17999999999999	74.98	Yellow

T 75.28 76.08000000000001 Red
Marker Name 26rs16891982
ASR 77.66 79.03
G 77.66 78.46000000000001 Blue
C 78.22999999999999 79.03 Yellow
Marker Name 27rs1335873
ASR 79.89 81.1
A 79.89 80.69000000000001 Green
T 80.3 81.10000000000001 Red
Marker Name 28rs1886510
ASR 80.52 82.01
G 80.52000000000001 81.31 Blue
A 81.17999999999999 82.01 Green
Marker Name 29rs730570
ASR 82.12 83.68
C 82.12 83.06 Yellow
T 82.67 83.67999999999999 Red
Marker Name 30rs5030240
ASR 85.05 86.71
G 85.05 85.67 Blue
A 86.05 86.71 Green
C 85.95 86.58 Yellow
Marker Name 31rs2304925
ASR 87.25 89.32
G 87.25 87.78999999999999 Blue
T 88.55 89.32000000000001 Red
Marker Name 32rs5997008
ASR 88.14 88.85
A 88.14 88.83 Green
C 88.23 88.85000000000001 Yellow
Marker Name 33rs3827760
ASR 90.46 91.43
G 90.46 91.02 Blue
A 90.73 91.42999999999999 Green
Marker Name 34rs2814778
ASR 91.78 93.27
C 91.78 92.77000000000001 Yellow
T 92.39 93.27 Red
Panel Name 34-PLEX
Marker Name 01rs1321333
ASR 32.62 35.85
C 32.620000000000005 33.75 Yellow
T 34.629999999999995 35.85 Red
Marker Name 02rs917118
ASR 31.54 34.26
G 31.540000000000003 32.75 Blue
A 33.46 34.26 Green
Marker Name 03rs1024116
ASR 31.87 34.99
G 31.870000000000005 33.06 Blue
A 33.77 34.99 Green
Marker Name 04rs7897550

ASR 36.44 39.3
C 36.440000000000005 37.27 Yellow
T 38.21 39.300000000000004 Red
Marker Name 05rs722098
ASR 36.94 39.87
G 36.940000000000005 37.74 Blue
A 39.07 39.87 Green
Marker Name 06rs10843344
ASR 40.55 43.09
C 40.550000000000004 41.35 Yellow
T 42.29 43.089999999999996 Red
Marker Name 07rs239031
ASR 41.21 44.01
C 41.21 42.01 Yellow
T 43.199999999999996 44.01 Red
Marker Name 08rs12913832
ASR 43.28 44.56
G 43.28 44.08 Blue
A 43.76 44.559999999999995 Green
Marker Name 09rs2040411
ASR 45.69 47.51
G 45.690000000000005 46.99 Blue
A 46.71 47.51 Green
Marker Name 10rs1978806
ASR 47.22 49.46
C 47.22 48.34 Yellow
T 48.660000000000004 49.46 Red
Marker Name 11rs773658
ASR 48.36 49.88
G 48.36 49.16 Blue
C 49.08 49.879999999999995 Yellow
Marker Name 12rs10141763
ASR 50.6 53.45
A 50.6 51.669999999999995 Green
T 51.91 53.45 Red
Marker Name 13rs182549
ASR 51.13 53.81
C 51.129999999999995 52.25 Yellow
T 53.01 53.809999999999995 Red
Marker Name 14rs1573020
ASR 52.36 54.19
G 52.36 53.86 Blue
A 52.94 54.19 Green
Marker Name 15rs896788
ASR 55.02 57.07
C 55.019999999999996 56.39 Yellow
T 56.220000000000006 57.07 Red
Marker Name 16rs2065160
ASR 56.85 58.04
G 56.849999999999994 57.75 Blue
A 57.24 58.04 Green
Marker Name 17rs2572307

ASR 58.03 59.66
G 58.03 58.83 Blue
A 58.86 59.66 Green
Marker Name 18rs2303798
ASR 59.78 61.12
C 59.78 60.58 Yellow
T 60.32 61.12 Red
Marker Name 19rs2065982
ASR 62.31 63.86
G 62.31 63.11 Blue
A 63.06 63.86 Green
Marker Name 20rs3785181
ASR 63.89 66.12
C 63.89 64.79 Yellow
T 65.32 66.12 Red
Marker Name 21rs881929
ASR 64.29 67.39
G 64.29 65.26 Blue
T 66.58999999999999 67.39 Red
Marker Name 22rs1498444
ASR 67.19 68.17
A 67.19 67.99000000000001 Green
C 67.36999999999999 68.17 Yellow
Marker Name 23rs1426654
ASR 70.48 72.13
C 70.47999999999999 71.28 Yellow
T 71.33 72.13000000000001 Red
Marker Name 24rs2026721
ASR 71.28 72.65
G 71.28 72.08000000000001 Blue
A 71.85 72.65 Green
Marker Name 25rs4540055
ASR 74.18 76.08
A 74.5 75.30000000000001 Green
C 74.17999999999999 74.98 Yellow
T 75.28 76.08000000000001 Red
Marker Name 26rs16891982
ASR 77.66 79.03
G 77.66 78.46000000000001 Blue
C 78.22999999999999 79.03 Yellow
Marker Name 27rs1335873
ASR 79.89 81.1
A 79.89 80.69000000000001 Green
T 80.3 81.10000000000001 Red
Marker Name 28rs1886510
ASR 80.52 82.01
G 80.52000000000001 81.31 Blue
A 81.17999999999999 82.01 Green
Marker Name 29rs730570
ASR 82.12 83.68
C 82.12 83.06 Yellow
T 82.67 83.67999999999999 Red

Marker Name 30rs5030240
ASR 85.05 86.71
G 85.05 85.67 Blue
A 86.05 86.71 Green
C 85.95 86.58 Yellow
Marker Name 31rs2304925
ASR 87.25 89.32
G 87.25 87.78999999999999 Blue
T 88.55 89.32000000000001 Red
Marker Name 32rs5997008
ASR 88.14 88.85
A 88.14 88.83 Green
C 88.23 88.85000000000001 Yellow
Marker Name 33rs3827760
ASR 90.46 91.43
G 90.46 91.02 Blue
A 90.73 91.42999999999999 Green
Marker Name 34rs2814778
ASR 91.78 93.27
C 91.78 92.77000000000001 Yellow
T 92.39 93.27 Red
Panel Name 34-PLEX Extra
Marker Name 02rs917118
ASR 69.9 71.5
G 69.89999999999999 70.55 Blue
A 70.80000000000001 71.5 Green
Marker Name 03rs1024116
ASR 32.4 34.99
G 32.400000000000006 33.24 Blue
A 33.77 34.99 Green
Marker Name 05rs722098
ASR 36.67 39.4
G 36.67 37.47 Blue
A 38.6 39.4 Green
Marker Name 06rs10843344
ASR 39.25 41.42
C 39.25 40.05 Yellow
T 40.620000000000005 41.42 Red
Marker Name 07rs239031
ASR 43.27 45.87
C 43.27 44.07 Yellow
T 45.06 45.870000000000005 Red
Marker Name 13rs182549
ASR 51.13 53.81
C 51.129999999999995 52.25 Yellow
T 53.01 53.809999999999995 Red
Marker Name 28rs1886510
ASR 80.52 82.01
G 80.52000000000001 81.31 Blue
A 81.17999999999999 82.01 Green
Panel Name 34-PLEX Extra Mod
Marker Name 02rs917118

ASR 69.9 71.5
G 69.89999999999999 70.55 Blue
A 70.80000000000001 71.5 Green
Marker Name 03rs1024116
ASR 32.4 34.99
G 32.400000000000006 33.24 Blue
A 33.77 34.99 Green
Marker Name 05rs722098
ASR 36.67 39.4
G 36.67 37.47 Blue
A 38.6 39.4 Green
Marker Name 06rs10843344
ASR 39.25 41.42
C 39.25 40.05 Yellow
T 40.620000000000005 41.42 Red
Marker Name 07rs239031
ASR 43.27 45.87
C 43.27 44.07 Yellow
T 45.06 45.870000000000005 Red
Marker Name 13rs182549
ASR 51.13 53.81
C 51.129999999999995 52.25 Yellow
T 53.01 53.809999999999995 Red
Marker Name 10rs1978806
ASR 47.0 49.9
C 47.0 48.0 Yellow
T 48.9 49.9 Red
Panel Name 34-Plex Electrophoretic Shift 2
Marker Name 01rs1321333
ASR 32.86 36.83
C 32.86 33.989999999999995 Yellow
T 35.61 36.830000000000005 Red
Marker Name 02rs917118
ASR 32.6 35.67
G 32.6 33.809999999999995 Blue
A 34.870000000000005 35.67 Green
Marker Name 03rs1024116
ASR 33.35 36.39
G 33.35 34.54 Blue
A 35.17 36.39 Green
Marker Name 04rs7897550
ASR 36.47 39.46
C 36.47 37.3 Yellow
T 38.37 39.46 Red
Marker Name 05rs722098
ASR 36.72 39.59
G 36.72 37.519999999999996 Blue
A 38.79 39.589999999999996 Green
Marker Name 06rs10843344
ASR 41.68 44.0
C 41.68 42.48 Yellow
T 43.2 44.0 Red

Marker Name 07rs239031
ASR 42.35 45.17
C 42.35 43.15 Yellow
T 44.36 45.17 Red
Marker Name 08rs12913832
ASR 44.16 45.64
G 44.160000000000004 44.96 Blue
A 44.84 45.64 Green
Marker Name 09rs2040411
ASR 46.48 48.49
G 46.480000000000004 47.78 Blue
A 47.690000000000005 48.49 Green
Marker Name 10rs1978806
ASR 47.82 49.26
C 47.82 48.940000000000005 Yellow
T 48.46 49.26 Red
Marker Name 11rs773658
ASR 48.55 50.04
G 48.550000000000004 49.35 Blue
C 49.24 50.04 Yellow
Marker Name 12rs10141763
ASR 50.32 52.76
A 50.32 51.38999999999999 Green
T 51.58 52.76 Red
Marker Name 13rs182549
ASR 51.0 53.29
C 51.0 52.120000000000005 Yellow
T 52.49 53.29 Red
Marker Name 14rs1573020
ASR 51.23 53.38
G 51.23 52.73 Blue
A 52.13 53.38 Green
Marker Name 15rs896788
ASR 54.01 56.51
C 54.01 55.38 Yellow
T 55.660000000000004 56.51 Red
Marker Name 16rs2065160
ASR 56.36 57.68
G 56.36 57.260000000000005 Blue
A 56.88 57.68 Green
Marker Name 17rs2572307
ASR 58.03 59.66
G 58.03 58.83 Blue
A 58.86 59.66 Green
Marker Name 18rs2303798
ASR 59.78 61.12
C 59.78 60.58 Yellow
T 60.32 61.12 Red
Marker Name 19rs2065982
ASR 62.31 63.86
G 62.31 63.11 Blue
A 63.06 63.86 Green

Marker Name 20rs3785181
ASR 63.89 66.12
C 63.89 64.79 Yellow
T 65.32 66.12 Red
Marker Name 21rs881929
ASR 64.29 67.39
G 64.29 65.26 Blue
T 66.58999999999999 67.39 Red
Marker Name 22rs1498444
ASR 67.19 68.17
A 67.19 67.99000000000001 Green
C 67.36999999999999 68.17 Yellow
Marker Name 23rs1426654
ASR 70.48 72.13
C 70.47999999999999 71.28 Yellow
T 71.33 72.13000000000001 Red
Marker Name 24rs2026721
ASR 71.28 72.65
G 71.28 72.08000000000001 Blue
A 71.85 72.65 Green
Marker Name 25rs4540055
ASR 74.18 76.08
A 74.5 75.30000000000001 Green
C 74.17999999999999 74.98 Yellow
T 75.28 76.08000000000001 Red
Marker Name 26rs16891982
ASR 77.66 79.03
G 77.66 78.46000000000001 Blue
C 78.22999999999999 79.03 Yellow
Marker Name 27rs1335873
ASR 79.89 81.1
A 79.89 80.69000000000001 Green
T 80.3 81.10000000000001 Red
Marker Name 28rs1886510
ASR 80.52 82.01
G 80.52000000000001 81.31 Blue
A 81.17999999999999 82.01 Green
Marker Name 29rs730570
ASR 82.12 83.68
C 82.12 83.06 Yellow
T 82.67 83.67999999999999 Red
Marker Name 30rs5030240
ASR 85.05 86.71
G 85.05 85.67 Blue
A 86.05 86.71 Green
C 85.95 86.58 Yellow
Marker Name 31rs2304925
ASR 86.42 88.52
G 86.42 86.96 Blue
T 87.75 88.52000000000001 Red
Marker Name 32rs5997008
ASR 88.14 88.85

A 88.14 88.83 Green
C 88.23 88.85000000000001 Yellow
Marker Name 33rs3827760
ASR 90.46 91.43
G 90.46 91.02 Blue
A 90.73 91.42999999999999 Green
Marker Name 34rs2814778
ASR 90.16 92.14
C 90.16 91.15 Yellow
T 91.26 92.14 Red
Panel Name 34-plex Elec Feb 2014
Marker Name 01rs1321333
ASR 32.62 35.85
C 32.620000000000005 33.75 Yellow
T 34.629999999999995 35.85 Red
Marker Name 02rs917118
ASR 31.54 34.26
G 31.540000000000003 32.75 Blue
A 33.46 34.26 Green
Marker Name 03rs1024116
ASR 31.87 34.99
G 31.870000000000005 33.06 Blue
A 33.77 34.99 Green
Marker Name 04rs7897550
ASR 36.44 39.3
C 36.440000000000005 37.27 Yellow
T 38.21 39.300000000000004 Red
Marker Name 05rs722098
ASR 36.94 39.87
G 36.940000000000005 37.74 Blue
A 39.07 39.87 Green
Marker Name 06rs10843344
ASR 40.55 43.09
C 40.550000000000004 41.35 Yellow
T 42.29 43.089999999999996 Red
Marker Name 07rs239031
ASR 41.21 44.01
C 41.21 42.01 Yellow
T 43.199999999999996 44.01 Red
Marker Name 08rs12913832
ASR 43.28 44.56
G 43.28 44.08 Blue
A 43.76 44.559999999999995 Green
Marker Name 09rs2040411
ASR 45.69 47.51
G 45.690000000000005 46.99 Blue
A 46.71 47.51 Green
Marker Name 10rs1978806
ASR 47.22 49.46
C 47.22 48.34 Yellow
T 48.660000000000004 49.46 Red
Marker Name 11rs773658

ASR 48.36 49.88
G 48.36 49.16 Blue
C 49.08 49.879999999999995 Yellow
Marker Name 12rs10141763
ASR 50.6 53.45
A 50.6 51.669999999999995 Green
T 51.91 53.45 Red
Marker Name 13rs182549
ASR 51.13 53.81
C 51.129999999999995 52.25 Yellow
T 53.01 53.809999999999995 Red
Marker Name 14rs1573020
ASR 52.36 54.19
G 52.36 53.86 Blue
A 52.94 54.19 Green
Marker Name 15rs896788
ASR 55.02 57.07
C 55.019999999999996 56.39 Yellow
T 56.220000000000006 57.07 Red
Marker Name 16rs2065160
ASR 56.85 58.04
G 56.849999999999994 57.75 Blue
A 57.24 58.04 Green
Marker Name 17rs2572307
ASR 58.03 59.66
G 58.03 58.83 Blue
A 58.86 59.66 Green
Marker Name 18rs2303798
ASR 59.78 61.12
C 59.78 60.58 Yellow
T 60.32 61.12 Red
Marker Name 19rs2065982
ASR 62.31 63.86
G 62.31 63.11 Blue
A 63.06 63.86 Green
Marker Name 20rs3785181
ASR 63.89 66.12
C 63.89 64.79 Yellow
T 65.32 66.12 Red
Marker Name 21rs881929
ASR 64.29 67.39
G 64.29 65.26 Blue
T 66.58999999999999 67.39 Red
Marker Name 22rs1498444
ASR 67.19 68.17
A 67.19 67.99000000000001 Green
C 67.36999999999999 68.17 Yellow
Marker Name 23rs1426654
ASR 70.48 72.13
C 70.47999999999999 71.28 Yellow
T 71.33 72.13000000000001 Red
Marker Name 24rs2026721

ASR 71.28 72.65
G 71.28 72.08000000000001 Blue
A 71.85 72.65 Green
Marker Name 25rs4540055
ASR 74.18 76.08
A 74.5 75.30000000000001 Green
C 74.17999999999999 74.98 Yellow
T 75.28 76.08000000000001 Red
Marker Name 26rs16891982
ASR 77.66 79.03
G 77.66 78.46000000000001 Blue
C 78.22999999999999 79.03 Yellow
Marker Name 27rs1335873
ASR 79.89 81.1
A 79.89 80.69000000000001 Green
T 80.3 81.10000000000001 Red
Marker Name 28rs1886510
ASR 80.52 82.01
G 80.52000000000001 81.31 Blue
A 81.17999999999999 82.01 Green
Marker Name 29rs730570
ASR 82.12 83.68
C 82.12 83.06 Yellow
T 82.67 83.67999999999999 Red
Marker Name 30rs5030240
ASR 85.05 86.71
G 85.05 85.67 Blue
A 86.05 86.71 Green
C 85.95 86.58 Yellow
Marker Name 31rs2304925
ASR 86.12 87.95
G 86.12 86.66 Blue
T 87.17999999999999 87.95 Red
Marker Name 32rs5997008
ASR 87.74 88.78
A 87.74 88.42999999999999 Green
C 88.16 88.78 Yellow
Marker Name 33rs3827760
ASR 90.46 91.29
G 90.46 91.02 Blue
A 90.59 91.28999999999999 Green
Marker Name 34rs2814778
ASR 89.64 91.85
C 89.64 90.63000000000001 Yellow
T 90.97 91.85 Red

Supplementary Files S2.8

Click here to download e-component: 34-PLEX_POP7_Panels.txt

Version GM v 3.0
Kit type: SNP
Chemistry Kit 34-PLEX none
Panel 34-Plex Electrophoretic Shift none
01rs1321333 - none
02rs917118 - none
03rs1024116 - none
04rs7897550 - none
05rs722098 - none
06rs10843344 - none
07rs239031 - none
08rs12913832 - none
09rs2040411 - none
10rs1978806 - none
11rs773658 - none
12rs10141763 - none
13rs182549 - none
14rs1573020 - none
15rs896788 - none
16rs2065160 - none
17rs2572307 - none
18rs2303798 - none
19rs2065982 - none
20rs3785181 - none
21rs881929 - none
22rs1498444 - none
23rs1426654 - none
24rs2026721 - none
25rs4540055 - none
26rs16891982 - none
27rs1335873 - none
28rs1886510 - none
29rs730570 - none
30rs5030240 - none
31rs2304925 - none
32rs5997008 - none
33rs3827760 - none
34rs2814778 - none
Panel 34-PLEX none
01rs1321333 - none
02rs917118 - none
03rs1024116 - none
04rs7897550 - none
05rs722098 - none
06rs10843344 - none
07rs239031 - none
08rs12913832 - none
09rs2040411 - none
10rs1978806 - none
11rs773658 - none
12rs10141763 - none
13rs182549 - none

14rs1573020 - none
15rs896788 - none
16rs2065160 - none
17rs2572307 - none
18rs2303798 - none
19rs2065982 - none
20rs3785181 - none
21rs881929 - none
22rs1498444 - none
23rs1426654 - none
24rs2026721 - none
25rs4540055 - none
26rs16891982 - none
27rs1335873 - none
28rs1886510 - none
29rs730570 - none
30rs5030240 - none
31rs2304925 - none
32rs5997008 - none
33rs3827760 - none
34rs2814778 - none
Panel 34-PLEX Extra none
02rs917118 - none
03rs1024116 - none
05rs722098 - none
06rs10843344 - none
07rs239031 - none
13rs182549 - none
28rs1886510 - none
Panel 34-PLEX Extra Mod none
02rs917118 - none
03rs1024116 - none
05rs722098 - none
06rs10843344 - none
07rs239031 - none
13rs182549 - none
10rs1978806 - none
Panel 34-Plex Electrophoretic Shift 2 none
01rs1321333 - none
02rs917118 - none
03rs1024116 - none
04rs7897550 - none
05rs722098 - none
06rs10843344 - none
07rs239031 - none
08rs12913832 - none
09rs2040411 - none
10rs1978806 - none
11rs773658 - none
12rs10141763 - none
13rs182549 - none
14rs1573020 - none

15rs896788 - none
16rs2065160 - none
17rs2572307 - none
18rs2303798 - none
19rs2065982 - none
20rs3785181 - none
21rs881929 - none
22rs1498444 - none
23rs1426654 - none
24rs2026721 - none
25rs4540055 - none
26rs16891982 - none
27rs1335873 - none
28rs1886510 - none
29rs730570 - none
30rs5030240 - none
31rs2304925 - none
32rs5997008 - none
33rs3827760 - none
34rs2814778 - none
Panel 34-plex Elec Feb 2014 none
01rs1321333 - none
02rs917118 - none
03rs1024116 - none
04rs7897550 - none
05rs722098 - none
06rs10843344 - none
07rs239031 - none
08rs12913832 - none
09rs2040411 - none
10rs1978806 - none
11rs773658 - none
12rs10141763 - none
13rs182549 - none
14rs1573020 - none
15rs896788 - none
16rs2065160 - none
17rs2572307 - none
18rs2303798 - none
19rs2065982 - none
20rs3785181 - none
21rs881929 - none
22rs1498444 - none
23rs1426654 - none
24rs2026721 - none
25rs4540055 - none
26rs16891982 - none
27rs1335873 - none
28rs1886510 - none
29rs730570 - none
30rs5030240 - none
31rs2304925 - none

32rs5997008 - none
33rs3827760 - none
34rs2814778 - none

Supplementary Files S2.5

Click here to download e-component: AIM-indelplex_POP7_bins.txt

Version GM v 3.0
Chemistry Kit AIM-indelplex
BinSet Name AIM-indelplex
Panel Name AIM-indelplex
Marker Name 1470
1 61.28 0.5 0.76
2 66.54 0.5 0.5
Marker Name 777
1 70.9 0.5 0.5
2 73.85 0.5 0.5
Marker Name 196
1 81.19 0.5 0.5
2 84.3 0.5 0.5
Marker Name 881
1 89.41 0.5 0.5
2 93.46 0.5 0.5
Marker Name 3122
1 98.52 0.5 0.5
2 103.05 0.5 0.5
Marker Name 548
1 107.97 0.5 0.5
2 109.91 0.5 0.5
3 104.91 0.5 0.5 mutant
Marker Name 659
1 116.64 0.5 0.63
2 118.92 0.5 0.5
Marker Name 2011
1 127.45 0.5 0.5
2 132.61 0.5 0.5
Marker Name 2929
1 150.73 0.5 0.5
2 152.95 0.5 0.5
Marker Name 593
1 156.67 0.5 0.5
2 158.85 0.5 0.5
Marker Name 798
1 168.67 0.5 0.5
2 174.57 0.5 0.5
Marker Name 1193
1 181.46 0.5 0.5
2 183.5 0.5 0.5
Marker Name 1871
1 191.3 0.5 0.5
2 193.25 0.5 0.66
Marker Name 17
1 200.28 0.5 0.5
2 204.2 0.5 0.5
Marker Name 2538
1 210.23 0.5 0.75
2 213.79 0.5 0.75
Marker Name 1644
1 222.95 0.5 0.5

2 224.81 0.5 0.5
Marker Name 3854
1 56.54 0.97 0.5
2 60.12 0.59 0.5
Marker Name 2275
1 70.9 0.5 0.5
2 77.92 0.5 0.5
Marker Name 3072
1 106.57 0.5 0.5
2 113.5 0.5 0.5
Marker Name 772
1 118.4 0.5 0.5
2 121.6 0.5 0.5
Marker Name 2313
1 128.4 0.5 0.5
2 137.67 0.5 0.5
Marker Name 397
1 164.01 0.5 0.5
2 168.03 0.5 0.5
Marker Name 1636
1 174.2 0.5 0.5
2 175.88 0.5 0.5
Marker Name 51
1 184.47 0.5 0.5
2 189.56 0.5 0.5
Marker Name 2431
1 209.32 0.5 0.75
2 213.694919173 0.5 0.75
Marker Name 2264
1 225.25 0.5 0.5
3 228.0 0.4 0.4 mutant
2 229.12 0.4 0.66
Marker Name 2256
1 58.21 0.5 0.5
2 61.37469617 4 0.5 0.5
Marker Name 128
1 68.29 0.5 0.5
2 71.41 0.5 0.5
Marker Name 15
1 78.8 0.5 0.5
2 89.72 0.5 0.64
Marker Name 2241
1 107.93 0.64 0.5
2 115.84 0.5 0.5
Marker Name 419
1 120.77 0.5 0.65
2 127.68 0.5 0.5
Marker Name 943
1 156.89 0.56 0.5
2 160.77 0.5 0.5
Marker Name 159
1 168.75 0.61 0.5

2	173.8	0.5	0.5	
Marker Name	2005			
1	184.78	0.5	0.55	
2	190.57	0.64	0.5	
Marker Name	250			
1	201.03	0.5	0.5	
2	203.1	0.5	0.5	
Marker Name	1802			
1	213.67	0.5	0.5	
2	216.64	0.5	0.5	
Marker Name	1607			
1	221.42	0.5	0.5	
2	223.03	0.62	0.5	
Marker Name	406			
1	66.96	0.5	0.5	
2	68.77	0.5	0.5	
Marker Name	1386			
1	72.66	0.5	0.71	
2	94.14	0.5	0.5	
Marker Name	1726			
1	105.6	0.5	0.5	
2	118.28	0.5	0.5	
Marker Name	3626			
1	142.49	0.5	0.5	
2	158.53	0.5	0.5	
Marker Name	360			
1	170.34	0.4	0.4	
3	171.2	0.4	0.4	mutant
2	172.15	0.4	0.4	
Marker Name	1603			
1	214.33	0.5	0.56	
2	218.0	0.5	0.5	
Marker Name	2719			
1	227.71	0.5	0.5	
2	231.5	0.5	0.5	
Marker Name	1734			
1	57.94	0.5	0.5	
2	61.77	0.5	0.5	
Marker Name	94			
1	91.31	0.5	0.5	
2	94.25	0.5	0.56	

Supplementary Files S2.6

Click here to download e-component: AIM-indelplex_POP7_Panels.txt

Version GM v 3.0
Kit type: MICROSATELLITE
Chemistry Kit AIM-indelplex none
Panel AIM-indelplex none

1470	blue	59.498930343000005	68.0	-	9	0.0	none	
777	blue	68.55843072	76.0	-	9	0.0	none	
196	blue	77.588872041	85.0	-	9	0.0	none	
881	blue	85.617309136	94.0	-	9	0.0	none	
3122	blue	95.320690594	103.8	-	9	0.0	none	
548	blue	104.2	110.8	-	9	0.0	none	
659	blue	113.556264981	120.0	-	9	0.0	none	
2011	blue	124.498164706	135.0	-	9	0.0	none	
2929	blue	146.955519673	153.8	-	9	0.0	none	
593	blue	154.7	160.0	-	9	0.0	none	
798	blue	165.32897286899998	176.0	-	9	0.0	none	
1193	blue	178.658800433	186.0	-	9	0.0	none	
1871	blue	188.823452509	195.5	-	9	0.0	none	
17	blue	197.084416798	205.0	-	9	0.0	none	
2538	blue	206.773228342	216.0	-	9	0.0	none	
1644	blue	219.483987195	226.0	-	9	0.0	none	
3854	Green	54.61523001	64.0	-	9	0.0	none	
2275	Green	68.507204408	80.0	-	9	0.0	none	
3072	Green	103.268780394	114.0	-	9	0.0	none	
772	Green	115.766532729	123.0	-	9	0.0	none	
2313	Green	125.024081548	139.5	-	9	0.0	none	
397	Green	161.189615282	170.0	-	9	0.0	none	
1636	Green	172.08629923200002	177.5	-	9	0.0	none	
51	Green	181.211398852	191.0	-	9	0.0	none	
2431	Green	206.293870962	216.0	-	9	0.0	none	
2264	Green	221.926816896	231.0	-	9	0.0	none	
2256	Yellow	56.680080642	64.0	-	9	0.0	none	
128	Yellow	66.0	72.0	-	9	0.0	none	
15	Yellow	76.5	91.0	-	9	0.0	none	
2241	Yellow	104.973806014	116.5	-	9	0.0	none	
419	Yellow	118.07762833899999	129.0	-	9	0.0	none	
943	Yellow	153.696434559	162.0	-	9	0.0	none	
159	Yellow	166.02409311300002	175.0	-	9	0.0	none	
2005	Yellow	181.501371715	192.0	-	9	0.0	none	
250	Yellow	198.065914853	204.0	-	9	0.0	none	
1802	Yellow	210.217316797	218.0	-	9	0.0	none	
1607	Yellow	218.9	224.0	-	9	0.0	none	
406	Red	64.678876207	69.5	-	9	0.0	none	
1386	Red	70.5	96.0	-	9	0.0	none	
1726	Red	104.0	121.0	-	9	0.0	none	
3626	Red	141.0	160.0	-	9	0.0	none	
360	Red	168.792202855	174.0	-	9	0.0	none	
1603	Red	213.0	219.0	-	9	0.0	none	
2719	Red	226.0	232.5	-	9	0.0	none	
1734	Red	56.285110841	63.5	-	9	0.0	none	
94	Green	88.363699841	96.0	-	9	0.0	none	

Supplementary File S3
Click here to download e-component: Supplementary Files S3.xlsx

Supplementary File S3. SNP and Indel genotypes used in the exercise as reference pop

File S3.1. PCA input file formatted for use with *Snipper* at: http://math

File S3.2. Training set file applicable to custom analyses (and used as fixed reference

File S3.3. Likelihood ratios from cross-validation of PCA input file genotypes (compa

Note: Worksheets S3.1 and S3.2 need to be placed in 'position 1' to use make the genotyp

LRs obtained with *Snipper* from complete profile data in Supplementary File S3.1

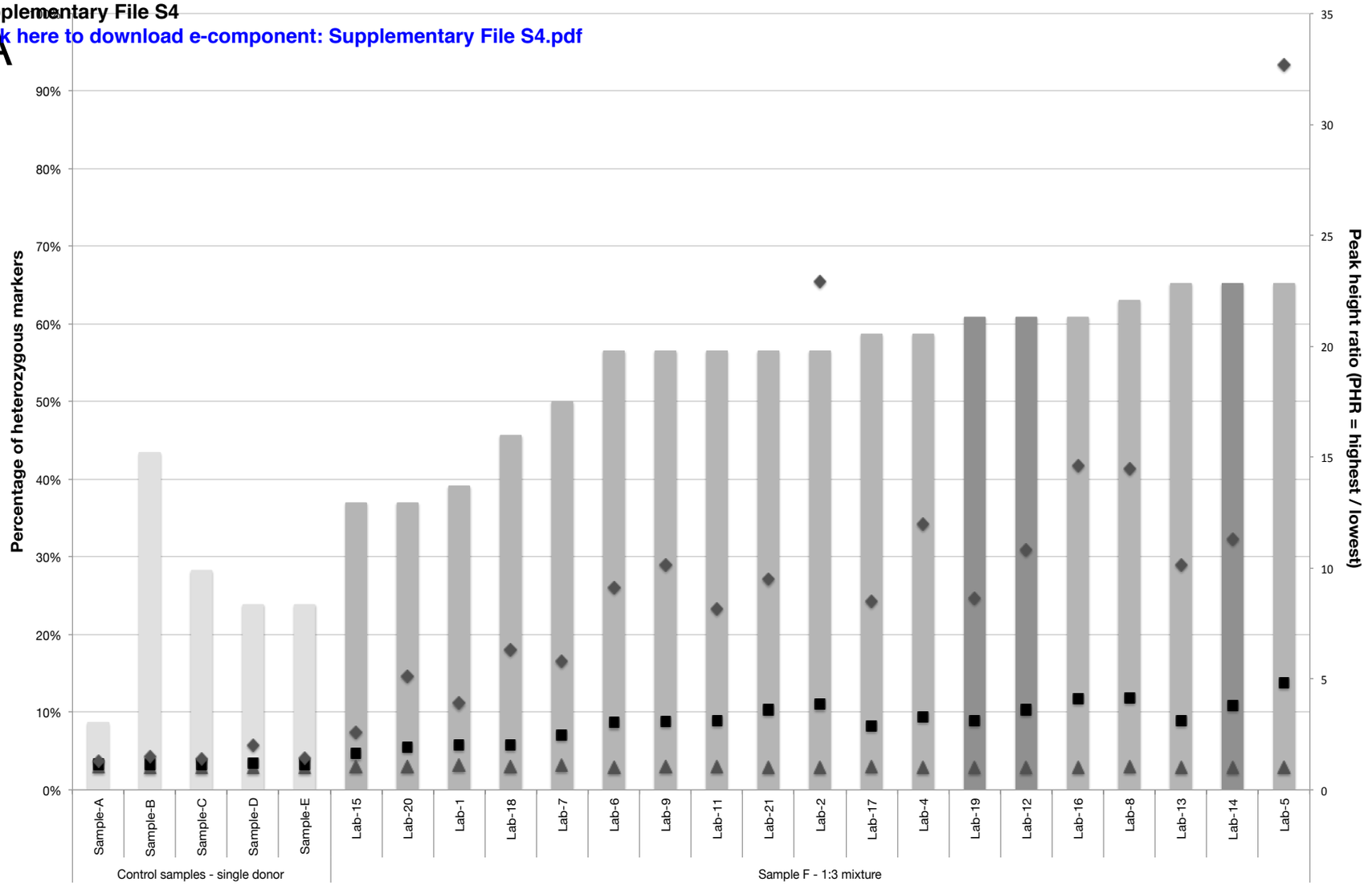
Inference	**34-plex, 3-group**
European	9947A is 2,118,840,589,047,061,020,672 times more likely EUROPE than E
East Asian	A is 361,148,635,069,545,024 times more likely EAST ASIA than EU
European	B is 64,191,487,284,485,608 times more likely EUROPE than EAST ASIA
East Asian	C is 13,115,706 times more likely EAST ASIA than AFRICA
East Asian	D is 248,539,593,557 times more likely EAST ASIA than EUROPE
African	E is 556,454,701,312,037,054,117,314,560 times more likely AFRIC

	46-plex, 4-group
European	9947A is 1,937,432,967,198 times more likely EUROPE than EAST ASIA
East Asian	A is 6,993,957 times more likely EAST ASIA than AMERICA
European	B is 143,659,679,122 times more likely EUROPE than EAST ASIA
East Asian	C is 131 times more likely EAST ASIA than EUROPE
American	D is 944,698,134 times more likely AMERICA than EAST ASIA
African	E is 3,229,841,442,838,053,650,432 times more likely AFRICA than

	80 Markers, 5-group
Oceanian	C is 153,747,536,542,653 times more likely OCEANIA than EAST ASIA

Click here to download e-component: Supplementary File S4.pdf

A



Supplementary File S4

A. Numbers of heterozygotes (bars) and PHR values (points) plotted for all participant's AIM-Indel data. Samples A-E are average values, sample F is per participant and includes laboratories #15, #20, #1 and #18 removed from the PHR comparisons made with the Kruskal-Wallis rank sum test summarized in C.

B

B. Grid of *p*-values for the pairwise comparison of numbers of heterozygotes in A-E (average number from 19 laboratories) and individual numbers per laboratory for sample F, applying a unilateral 2-sample test for equality-of-proportions (with continuity correction). Grey cells mark significant *p*-values.

	Sample-A	Sample-B	Sample-C	Sample-D	Sample-E	Lab-15	Lab-20	Lab-1	Lab-18	Lab-7	Lab-6	Lab-9	Lab-11	Lab-21	Lab-2	Lab-17	Lab-4	Lab-19	Lab-12	Lab-16	Lab-8	Lab-13	Lab-14
Sample-B	0.00018																						
Sample-C	0.01582	0.90390																					
Sample-D	0.04519	0.96120	0.59380																				
Sample-E	0.04519	0.96120	0.59380	0.50000																			
Lab-15	0.00144	0.66470	0.25230	0.12860	0.12860																		
Lab-20	0.00144	0.66470	0.25230	0.12860	0.12860	0.50000																	
Lab-1	0.00074	0.58380	0.18880	0.08909	0.08909	0.50000	0.50000																
Lab-18	8.85e-05	0.50000	0.06527	0.02441	0.02441	0.26260	0.26260	0.33650															
Lab-7	1.88e-05	0.33800	0.02726	0.00875	0.00875	0.14650	0.14650	0.20070	0.41730														
Lab-6	1.50e-06	0.14860	0.00568	0.00146	0.00146	0.04729	0.04729	0.07201	0.20210	0.33800													
Lab-9	1.50e-06	0.14860	0.00568	0.00146	0.00146	0.04729	0.04729	0.07201	0.20210	0.33800	0.50000												
Lab-11	1.50e-06	0.14860	0.00568	0.00146	0.00146	0.04729	0.04729	0.07201	0.20210	0.33800	0.50000	0.50000											
Lab-21	1.50e-06	0.14860	0.00568	0.00146	0.00146	0.04729	0.04729	0.07201	0.20210	0.33800	0.50000	0.50000	0.50000										
Lab-2	1.50e-06	0.14860	0.00568	0.00146	0.00146	0.04729	0.04729	0.07201	0.20210	0.33800	0.50000	0.50000	0.50000	0.50000									
Lab-17	6.09e-07	0.10540	0.00313	0.00075	0.00075	0.03016	0.03016	0.04761	0.14830	0.33800	0.50000	0.50000	0.50000	0.50000	0.50000								
Lab-4	6.09e-07	0.10540	0.00313	0.00075	0.00075	0.03016	0.03016	0.04761	0.14830	0.33800	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000							
Lab-19	2.39e-07	0.07201	0.00166	0.00037	0.00037	0.01851	0.01851	0.03028	0.10500	0.20070	0.41620	0.41620	0.41620	0.41620	0.41620	0.50000	0.50000						
Lab-12	2.39e-07	0.07201	0.00166	0.00037	0.00037	0.01851	0.01851	0.03028	0.10500	0.20070	0.41620	0.41620	0.41620	0.41620	0.41620	0.50000	0.50000	0.50000					
Lab-16	2.39e-07	0.07201	0.00166	0.00037	0.00037	0.01851	0.01851	0.03028	0.10500	0.20070	0.41620	0.41620	0.41620	0.41620	0.41620	0.50000	0.50000	0.50000	0.50000				
Lab-8	9.09e-08	0.04729	0.00085	0.00017	0.00017	0.01091	0.01091	0.01851	0.07144	0.14650	0.33530	0.33530	0.33530	0.33530	0.33530	0.41540	0.41540	0.50000	0.50000	0.50000			
Lab-13	3.34e-08	0.02980	0.00041	7.98e-05	7.98e-05	0.00616	0.00616	0.01084	0.04667	0.10280	0.26080	0.26080	0.26080	0.26080	0.26080	0.33380	0.33380	0.41450	0.41450	0.41450	0.50000		
Lab-14	3.34e-08	0.02980	0.00041	7.98e-05	7.98e-05	0.00616	0.00616	0.01084	0.04667	0.10280	0.26080	0.26080	0.26080	0.26080	0.26080	0.33380	0.33380	0.41450	0.41450	0.41450	0.50000	0.50000	
Lab-5	3.335e-08	0.02980	0.00041	7.98e-05	7.98e-05	0.006161	0.006161	0.01084	0.04667	0.10280	0.26080	0.26080	0.26080	0.26080	0.26080	0.33380	0.33380	0.41450	0.41450	0.41450	0.50000	0.50000	0.50000

C **c.** Grid of *p*-values for pairwise comparisons of PHR values applying a Kruskal-Wallis rank sum test (grey cells: significant values)

	Sample-A	Sample-B	Sample-C	Sample-D	Sample-E	Lab-7	Lab-6	Lab-9	Lab-11	Lab-21	Lab-2	Lab-17	Lab-4	Lab-19	Lab-12	Lab-16	Lab-8	Lab-13	Lab-14
Sample-B	0.24530																		
Sample-C	0.36500	0.05085																	
Sample-D	0.60150	0.26500	0.40090																
Sample-E	0.36080	0.86880	0.31060	0.41180															
Lab-7	0.00777	1.61e-06	3.58e-05	0.00033	5.56e-05														
Lab-6	0.01237	2.36e-06	9.51e-05	0.00033	8.82e-05	0.33620													
Lab-9	0.00341	1.92e-07	3.36e-06	5.80e-05	1.15e-05	0.37810	0.89810												
Lab-11	0.00727	6.92e-07	2.04e-05	0.00017	3.78e-05	0.17310	0.88360	0.67380											
Lab-21	0.01237	2.93e-06	0.00011	0.00026	8.82e-05	0.05445	0.38970	0.30540	0.36020										
Lab-2	0.00341	1.05e-07	3.36e-06	5.80e-05	1.15e-05	0.13820	0.71430	0.49830	0.78370	0.54590									
Lab-17	0.00672	5.98e-07	1.58e-05	0.00018	3.07e-05	0.29770	0.90090	1.00000	0.73530	0.24030	0.60590								
Lab-4	0.01128	1.29e-06	3.01e-05	0.00023	8.05e-05	0.28880	0.97160	0.88680	0.76230	0.34570	0.70870	0.86950							
Lab-19	0.00524	2.40e-07	1.24e-05	8.41e-05	2.52e-05	0.12990	0.87620	0.65260	0.86260	0.44620	1.00000	0.56700	0.74910						
Lab-12	0.01670	1.23e-06	2.64e-05	0.00020	8.41e-05	0.29780	0.86260	0.71620	0.97240	0.57960	0.88980	0.78760	0.89290	0.98040					
Lab-16	0.01034	8.91e-07	2.06e-05	0.00012	6.48e-05	0.08159	0.43600	0.38670	0.42580	0.98620	0.60350	0.32880	0.41900	0.55520	0.62300				
Lab-8	0.00582	3.67e-07	9.82e-06	9.93e-05	2.08e-05	0.07849	0.43810	0.31990	0.50010	0.94620	0.64900	0.32120	0.45560	0.58730	0.62070	0.97450			
Lab-13	0.00882	6.68e-07	1.14e-05	0.00017	4.31e-05	0.55370	0.81810	0.84370	0.64550	0.30840	0.46980	0.81050	0.71320	0.57540	0.58600	0.37510	0.28170		
Lab-14	0.00276	3.69e-08	1.49e-06	2.58e-05	5.86e-06	0.13170	0.53240	0.41140	0.63380	0.78000	0.88250	0.47200	0.63160	0.68580	0.72040	0.80340	0.87950	0.34400	
Lab-5	0.00544	2.62e-07	7.86e-06	8.05e-05	1.74e-05	0.11850	0.59910	0.45970	0.74250	0.69340	0.96070	0.51230	0.64300	0.75560	0.68580	0.82750	0.91550	0.38310	0.92930

Supplementary File S5 Next generation sequencing experiments using exercise PCR multiplexes.

1. Post-PCR DNA processing for SNP analysis with the MiSeq

Libraries were prepared directly from PCR products using the Illumina TruSeq ChIP sample preparation kit. Libraries were then run on the MiSeq with the 300 cycle version 2 reagent kit and sequences aligned to a custom 'genome' containing the reference sequences for all 34 SNPs in a single unified strand using Burrows-Wheeler alignment. SNP genotypes were called using GATK.

2. Post-PCR DNA processing for Indel and SNP analysis with the TFS-LT Ion PGM™

Libraries were prepared directly from PCR products using the TFS-LT Ion Xpress™ Plus gDNA fragment library preparation protocol applying the Ion Xpress™ Plus Fragment Library Kit. This kit processes DNA not amplified using AmpliSeq primers by enabling ligation and nick repair reactions. Thereafter sequencing followed standard protocols using kits: Ion OneTouch™ 200 Template v2 and Ion PGM™ Sequencing 200 v2. Sequences were aligned to custom BED files and genotypes called from human genome build hg19 using TFS-LT Torrent Suite™ 4.0.2.

3. Genotyping performance for samples A-E

Both NGS systems gave comparable high levels of genotyping performance for SNP analysis. Ion PGM™ had just singleton no-calls or missing data (no sequences detected carrying expected SNP sites), while MiSeq gave the only miscalled genotype in rs5030240, where a sequence ratio of A=5810 / G=9866 was recorded as a GG, although it was detected as atypical.

Indel genotyping performance with the Ion PGM™ was slightly lower, but this could be due to alignment issues. The Indel rs60612424 (MID-3854) was not detected in any samples, while C had a disproportionately high number of no-calls that might be the result of population-specific flanking indels blocking secure alignment to the reference sequence.

4. Summary tables of genotyping performance from SNP analysis with both NGS systems and Indel analysis with Ion PGM™. Matches count concordant genotype calls made in both CE and NGS.

Ion PGM™ sequence data for 34 SNPs

	A	B	C	D	E	F
NGS no-calls	1	0	0	1	1	1
SNaPshot no-calls	1	0	0	0	0	0
Ion PGM miscalls	0	0	0	0	0	4
Missing data	0	1	1	1	0	1
NGS-CE genotype matches *	32	33	33	32	33	28
Heterozygote number	10	5	10	9	10	14

MiSeq sequence data for 34 SNPs

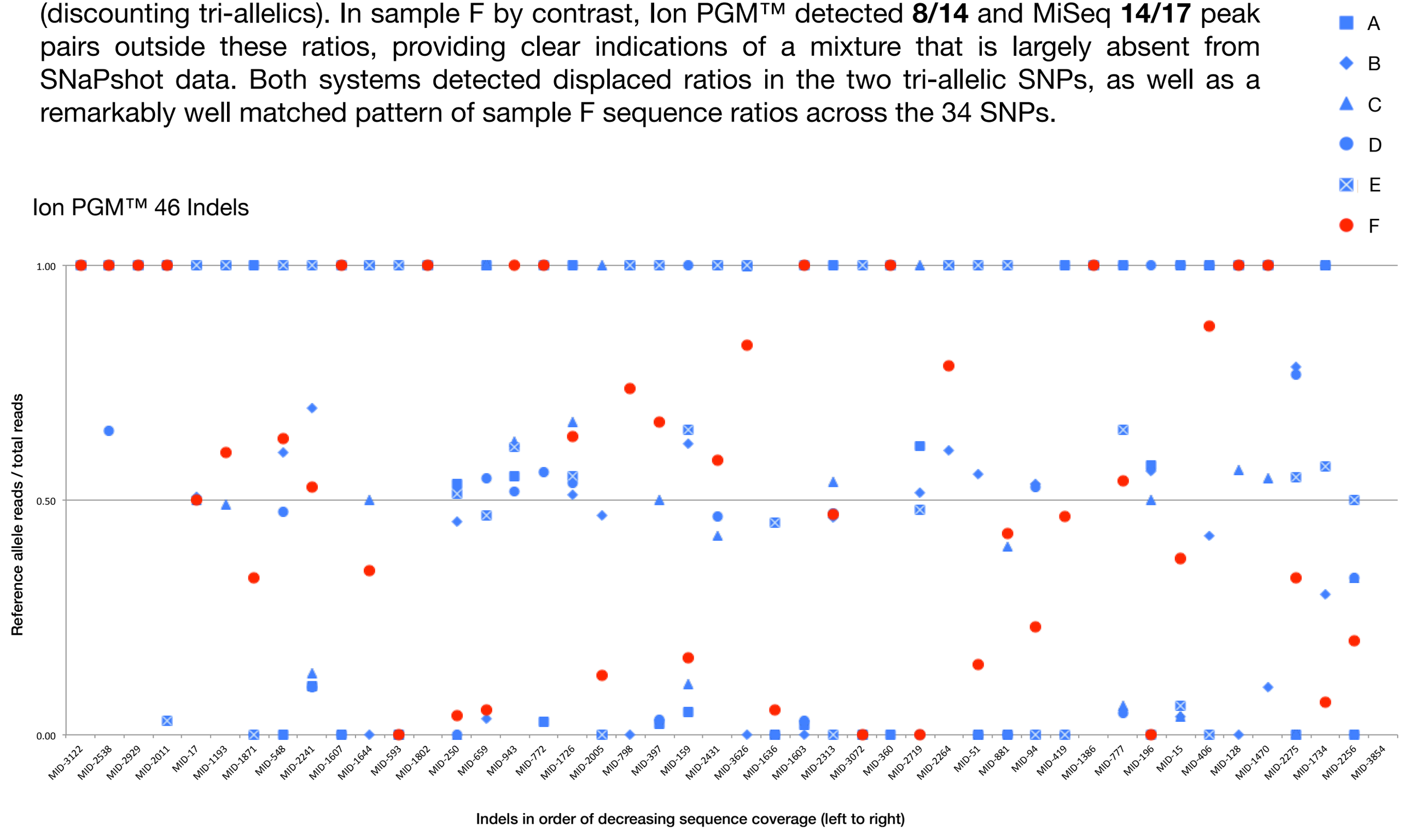
A	B	C	D	E	F
0	0	0	0	0	0
1	0	0	0	0	0
0	0	1	0	0	4
0	0	0	0	0	0
33	34	33	34	34	30
10	5	9	10	10	17

Ion PGM™ sequence data for 46 Indels

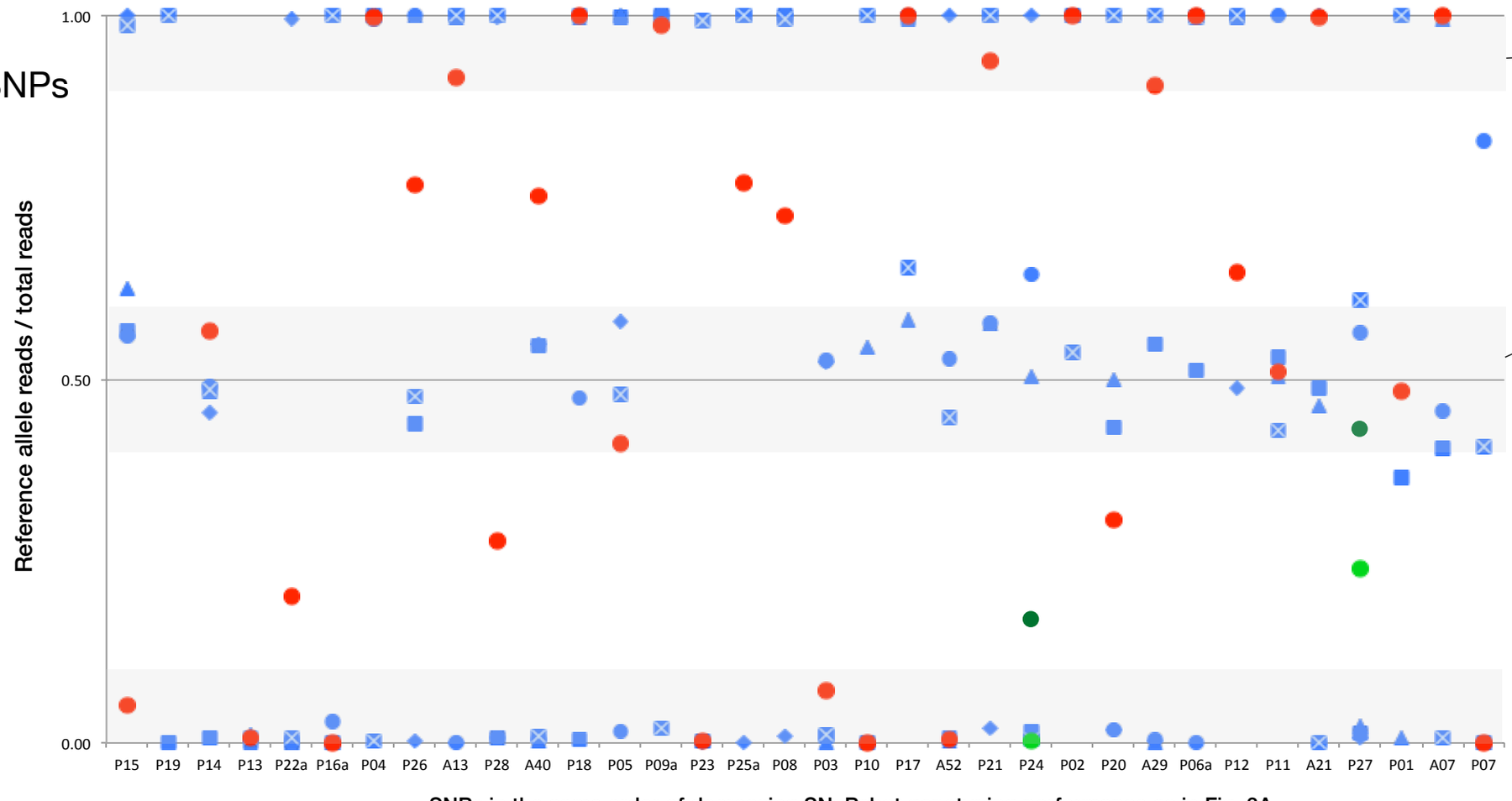
	A	B	C	D	E	F
Ion PGM no-calls	4	1	12	3	1	3
PCR-to-CE no-calls	0	0	0	0	0	0
Ion PGM miscalls	0	3	3	0	0	6
Missing data	1	2	2	1	1	1
NGS-CE genotype matches *	41	40	29	42	44	36
Heterozygote number	4	17	11	10	10	21

* Matching genotypes denote calls concordant with conventional CE analysis

5. NGS analysis of mixed sample F indicated a higher number of displaced sequence ratios positioned outside ranges around an ideal midline ratio of 0.5:0.5, although Indel genotyping showed several irregular ratios for A-E likely due to alignment issues. SNP genotyping with both platforms was very sensitive to imbalanced sequence ratios in F. In samples A-E, Ion PGM™ detected **4/44** peak pairs outside nominal 0.4:0.6-0.5:0.5 sequence ratio ranges, and MiSeq **2/44** (discounting tri-allelics). In sample F by contrast, Ion PGM™ detected **8/14** and MiSeq **14/17** peak pairs outside these ratios, providing clear indications of a mixture that is largely absent from SNaPshot data. Both systems detected displaced ratios in the two tri-allelic SNPs, as well as a remarkably well matched pattern of sample F sequence ratios across the 34 SNPs.



Ion PGM™ 34 SNPs



Nominal homozygote ratios for normal DNA of 0.1:0.9 - 0:1

Nominal heterozygote ratios 0.4:0.6 - 0.5:0.5

A

B

C

D

E

F

F: Two ratios of tri-allelic SNPs

MiSeq 34 SNPs